



Elaboration de ressources électroniques pour les noms composés de type N (E+DET=G) N=G du grec moderne

Anthoula Kyriakopoulou

► To cite this version:

Anthoula Kyriakopoulou. Elaboration de ressources électroniques pour les noms composés de type N (E+DET=G) N=G du grec moderne. Linguistique. Université Paris-Est, 2011. Français. NNT : 2011PEST1029 . pastel-00666189

HAL Id: pastel-00666189

<https://pastel.archives-ouvertes.fr/pastel-00666189>

Submitted on 3 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Paris-Est
Institut Gaspard-Monge

Ecole Doctorale Information, Communication, Modélisation et Simulation
Laboratoire d'Informatique

THÈSE

Pour obtenir le grade de
Docteur de l'Université de Paris-Est

Discipline : Informatique Linguistique

par
Anthoula Kyriakopoulou

le 25 mars 2011

Les noms composés de type $N (E+DET:G) N:G$ en grec
moderne

Directeur de thèse : Panayota-Tita Kyriakopoulou

Jury

Anna ANASTASSIADIS-SYMEONIDIS (rapporteur)

Denis MAUREL (rapporteur)

Panayota-Tita KYRIACOPOULOU

Eric LAPORTE

Aggeliki FOTOPOULOU

Sébastien PAUMIER

© UPE

N° attribué par la bibliothèque
10101M1L1V111111

A Tita's Team

Remerciements

Dans le cadre de cette thèse, j'ai eu l'occasion de collaborer et de partager des moments créatifs avec plusieurs personnes. Que toutes ces personnes soient remerciées ici :

Tita Kyriacopoulou pour avoir dirigé cette thèse, pour m'avoir donné les moyens pour l'effectuer, pour sa confiance, ses conseils et son soutien tout au long de ce travail.

Eric Laporte pour l'accueil au labo, pour ses conseils et pour avoir accepté d'être membre de mon jury.

Denis Maurel et Anna Anastassiadis-Syméonidis pour leurs corrections et suggestions bibliographiques, pour leurs encouragements et pour avoir accepté de rapporter cette thèse.

Aggeliki Fotopoulou pour sa collaboration précieuse, ses suggestions bibliographiques, ses encouragements et pour avoir accepté d'être membre de mon jury.

Sébastien Paumier pour sa disponibilité et ses remarques et pour avoir accepté de participer à mon jury.

Béatrice Lamiroy pour ses remarques et ses suggestions bibliographiques ainsi que pour le matériel qu'elle a mis à ma disposition.

Stavroula Voyatzi pour son investissement personnel dans mon travail, pour la relecture minutieuse de ce mémoire, pour ses suggestions et ses remarques toujours pertinentes et pour les discussions animées tout au long de ce travail. Mais surtout pour sa générosité et son amitié.

Maïté Briselet, Brigitte O'rorke, Estelle Maillebau et Emilie Guimier de Neef pour leur accueil chaleureux à Lannion et pour le travail en équipe dans le cadre des projets TRANSAT et pagesjaunes.fr de France Telecom.

Tous les membres de l'équipe informatique linguistique de l'IGM pour leurs remarques et les discussions intéressantes lors des réunions des doctorants.

Tous les membres de l'équipe informatique linguistique de l'Université Aristote de Thessaloniki, et notamment Ionas Michailidis, Rania Voskaki, Olympia Tsaknaki, Kelly Ioannidou et Vassiliki Foufi, pour leur persistance, leur collaboration et leur disponibilité.

Takuya Nakamura pour les discussions (linguistiques et autres), pour le soutien moral, pour sa franchise et sa disponibilité, notamment lors de mes recherches bibliographiques.

Jean-Pierre Heymann pour la relecture d'une partie du mémoire.

Patrice Herault pour l'assistance technique.

Pierre Guillon, Guillaume Blin, Damien Masson et Matthieu Constant pour leur soutien et leur présence rassurante.

Sylvie Cach pour sa disponibilité et son aide au niveau administratif.

Les copains doctorants et docteurs André Esteves Oliveira, Eli Murguia, Nikolas Sklavounos, Kostas Rokas, Konstantinos Kritsis, Lamprini Kolovou, Alessandra Savelli, Marianna Mini pour la solidarité du même parcours choisi.

Les amis Mimis Andreadis, Lola Gantsidou, Xenia Karamouza, Soula Biglikoudi, Christina Papathanassi, Evangelia Bouloubassi, Kelly Papadaki, Laura Crognale, Yanna Karypi, Per Wennlund, Leonidas Koutsos et Christina Parameriti pour avoir partagé les moments difficiles de ce parcours et pour leur passage à Paris.

Yannis et Alexis Souriadakis pour les moments de joie lors de mon séjour à Paris.

Mes parents.

Tous ceux qui ont partagé leurs connaissances et leur expérience avec moi.

Alexis et Luca Crognale, pour leur présence radieuse dans ma vie.

Résumé

L'objectif de cette recherche est la construction manuelle de ressources lexicales pour les noms composés grecs qui sont définis par la structure morphosyntaxique : *Nom (E+Déterminant au génitif) Nom au génitif*, notés $N(E+DET:G)N:G$ (e.g. ζώνη ασφαλείας/ceinture de sécurité). Les ressources élaborées peuvent être utilisées pour leur reconnaissance lexicale automatique dans les textes écrits et dans d'autres applications du TAL.

Notre travail s'inscrit dans la perspective de l'élaboration du lexique-grammaire général du grec moderne en vue de l'analyse automatique des textes écrits. Le cadre théorique et méthodologique de cette étude est celui du lexique-grammaire (M. Gross 1975, 1977), qui s'appuie sur la grammaire transformationnelle harisienne.

Notre travail s'organise en cinq parties. Dans la première partie, nous délimitons l'objet de notre travail tout en essayant de définir la notion fondamentale qui régit notre étude, à savoir celle de figement. Dans la deuxième partie, nous présentons la méthodologie utilisée pour le recensement de nos données lexicales et nous étudions les phénomènes de variation observés au sein des noms composés de type $N(E+DET:G)N:G$. La troisième partie est consacrée à la présentation des différentes sous-catégories des $N(E+DET:G)N:G$ identifiées lors de l'étape du recensement et à l'étude de leur structure lexicale interne. La quatrième partie porte sur l'étude syntaxico-sémantique des $N(E+DET:G)N:G$. Enfin, dans la cinquième partie, nous présentons les différentes méthodes de représentation formalisée que nous proposons pour nos données lexicales en vue de leur reconnaissance lexicale automatique dans les textes écrits.

Des échantillons représentatifs des ressources élaborées sont présentés en Annexe.

Mots-clés : traitement automatique des langues (TAL), approche linguistique, nom composé, unité lexicale multi-mots, syntagme nominal semi-figé, méthodes de recensement, variation, traits sémantiques, dictionnaire électronique morphologique, grammaire locale, graphe d'automates à états finis, table du lexique-grammaire

Abstract

The object of this research is the manual construction of lexical resources for the Greek compound nouns defined by the following morphosyntactic structure: *Noun (E+Determiner in genitive) Noun in genitive*, $(N (E+DET:G) N:G)$ (e.g. ζώνη ασφαλείας/safety belt). The elaborated resources may be used for their automatic recognition in written texts and other NLP applications.

Our study is part of the general lexicon-grammar for Modern Greek in view of automatic processing of written texts. Our theoretical and methodological framework is that of lexicon-grammar (M. Gross 1975, 1977), based on the Transformational Grammar principles defined by Z. S. Harris.

Our study is organised into five parts. In the first part, we give an overview of the core notion governing our research: the notions of (fixed) multiword expression (MWE). In the second part, we present the methodology used to collect our lexical data and we study the variation phenomena observed within the framework of the $N (E+DET:G) N:G$. The third part is dedicated to the presentation of the different $N (E+DET:G) N:G$ categories identified in the listing phase and to the study of their lexical composition. The fourth concerns the syntactical and semantic study of the $N (E+DET:G) N:G$. Finally, the fifth part deals with the formal representation methods we propose for our lexical data in view of their lexical recognition in Greek written texts.

Representative samples of the elaborated resources are illustrated in Appendix.

Keywords: natural language processing (NLP), linguistic approach, compound noun, multiword lexical unit, semi-fixed noun group, listing methods, semantic description, electronic morphological dictionary, local grammar, finite state automata, lexicon-grammar table

Table des matières

Introduction	18
REMIERE PARTIE : DELIMITATION DE L'ETUDE	24
Chapitre 1. Travaux antérieurs	25
1.1 Survol de la littérature sur les noms composés en grec moderne.....	25
1.1.1 Classification des noms composés en grec moderne	37
1.2 Survol de la littérature sur les noms composés en français.....	39
Chapitre 2. La notion de figement.....	42
2.1 Introduction	43
2.2 Les critères généraux et les propriétés générales du phénomène de figement.....	44
2.2.1 La polylexicalité.....	45
2.2.2 La non-compositionnalité du sens (ou les contraintes sémantiques)	46
2.2.3 Le blocage des propriétés transformationnelles	48
2.2.4 La non-actualisation des éléments constitutifs.....	50
2.2.5 La non-prédication des noms composés.....	52
2.2.6 Le blocage des paradigmes synonymiques (ou les contraintes lexicales).....	53
2.2.7 La non-insertion	56
2.2.8 La coordination	58
2.2.9 Le degré de figement.....	59
2.2.10 Le défigement.....	62
2.2.11 Etymologie	63
2.3 Distinction des $N(E + DET:G) N:G$ par rapport au degré de leur figement	66
2.3.1 Les $N(E + DET:G) N:G$ entièrement figés	66
2.3.2 Les $N(E + DET:G) N:G$ semi-figés (ou productifs)	69
2.3.3 Conclusion.....	70
DEUXIEME PARTIE : LE RECENSEMENT DES DONNEES.....	72
Chapitre 3. Brève description du dictionnaire électronique morphologique	68
3.1 Introduction	73

3.2	Le dictionnaire électronique morphologique des noms composés du grec moderne	74
Chapitre 4. Le recensement..... 74		
4.1	Introduction	79
4.2	Méthodes de recensement	80
4.2.1	Validation des entrées $N (E + DET:G) N:G$ du DELAC existant	80
4.2.2	Méthode manuelle : compilation de grammaires et de dictionnaires d'usage	89
4.2.3	Méthode semi-automatique : application de motifs morphosyntaxiques à des corpus de textes écrits	102
4.2.4	Proposition de sous-catégorisation des $N (E + DET:G) N:G$	110
Chapitre 5. La variation..... 109		
5.1	Introduction	114
5.2	Les noms composés binaires	116
5.2.1	Variantes lexicales.....	116
5.2.1.1	Possibilité de substitution des formes savantes par des formes démotiques	116
5.2.1.2	Possibilité de substitution des formes du grec par des emprunts	120
5.2.1.3	Possibilité de substitution par des formes en latin	122
5.2.1.4	Possibilité de substitution du nom-tête par son diminutif.....	123
5.2.1.5	Possibilité de substitution du deuxième composant nominal par une forme non-lexicale	123
5.2.1.6	Variation de nombre.....	123
5.2.2	Variantes orthographiques.....	127
5.2.3	Possibilité de substitution par des formes abrégées	128
5.2.4	Variantes phonologiques	129
5.2.5	Variantes syntaxiques.....	132
5.2.5.1	La réduction du déterminant	132
5.2.5.2	La réduction du nom-tête	133
5.2.5.3	La réduction du deuxième composant nominal.....	134
5.2.5.4	La permutation	135
5.2.5.5	Les variantes de type AN	136
5.2.5.6	Les variantes soudées	138
5.2.5.7	Les variantes en $N Prép (E+DET) N$	140
5.2.5.8	Les $N (E+Prép) (E+DET:G) N:G$	141

5.3 Les noms composés complexes (ou surcomposés) de structure $N1 (E+DET:G) A:G N2:G$	141
5.3.1 Variantes lexicales.....	142
5.3.1.1 Possibilité de substitution des formes savantes par des formes démotiques 142	
5.3.1.2 Possibilité de substitution d'un mot par un autre mot du même registre ...	144
5.3.1.3 Possibilité de substitution du singulier par le pluriel	145
5.3.2 Variantes orthographiques.....	146
5.3.3 Possibilité de substitution par des formes abrégées	146
5.3.3 Variantes syntaxiques.....	147
TROISIEME PARTIE : STRUCTURE LEXICALE INTERNE	154
Chapitre 6. Les différentes catégories et sous-catégories des $N (E + DET:G) N:G$	150
6.1 Introduction	155
6.1 Les noms composés binaires de structure $N (E+DET:G) N:G$	156
6.3 Les noms composés complexes (ou surcomposés)	156
6.4 Les noms composés entièrement figés	160
6.5 Les syntagmes nominaux semi-figés (ou productifs).....	161
6.6 Les entités nommées	163
6.7 Les termes de spécialité.....	167
6.8 Les syntagmes nominaux figés faisant partie d'une phrase simple figée et qui ont également un emploi autonome en dehors de celle-ci.....	168
Chapitre 7. La structure lexicale interne des noms composés $N (E+DET:G) N:G$	166
7.1 Introduction	171
7.2 Les déterminants	171
7.2.1 Les déterminants dans les noms composés de structure $N (E+DET:G) N:G$	171
7.2.1.1 Déterminants définis	172
7.2.1.1.1 Articles définis.....	172
7.2.1.1.1 Possessifs sans source.....	174
7.2.1.2 Déterminants indéfinis	176
7.2.1.2.1 Articles et adjectifs indéfinis	176
7.2.1.2.2 Déterminants numéraux.....	177
7.2.1.2.3 Déterminant zéro	178
QUATRIEME PARTIE : DESCRIPTION LINGUISTIQUE	182
Chapitre 8. Etude syntaxico-sémantique des $N (E+DET:G) N:G$	178

8.1	Introduction	183
8.2	Le génitif dans les $N (E+DET:G) N:G$	184
8.2.1	Le système casuel du grec moderne	184
8.2.2	La classification du génitif adnominal en grec moderne.....	185
8.2.3	La formation du complément de nom au génitif en grec moderne	188
8.2.4	Blocage des dérivations syntaxiques régulières et reproductibles dans les $N (E+DET:G) N:G$	191
Chapitre 9. Distinction entre $N (E+DET:G) N:G$ prédicatifs et $N (E+DET:G) N:G$ non-prédicatifs		187
9.1	Délimitation de l'étude des noms prédicatifs.....	192
9.2	La notion de nom prédicatif et de verbe support.....	192
9.3	Les critères syntaxiques de distinction entre noms prédicatifs et noms non-prédicatifs	194
9.3.1	Le sujet sémantique du nom prédicatif est le sujet grammatical du verbe support	195
9.3.2	Existence d'un groupe nominal indépendant qui soit synonyme de la phrase à verbe support et qui conserve le complément <i>Prép NI</i>	197
9.3.3	La question par <i>τι;/que ?</i> , souvent douteuse.....	198
9.3.4	Descente ou montée de l'adverbe.....	200
9.4	Problèmes de distinction entre expression à éléments figés et expression à verbe support.....	202
Chapitre 10. Classification sémantique des $N (E + DET:G) N:G$ et des $NI (E+DET:G) A:G N2:G$		199
10.1	Introduction	204
10.2	Les traits sémantiques élémentaires	205
10.3	Les conventions du codage	206
10.3.1	Traits binaires	206
10.3.2	Traits exclusifs	206
10.4	Les $N (E+DET:G) N:G$ et les $NI (E+DET:G) A:G N2:G$ humains	208
10.4.1	La notion sémantique « humain »	208
10.4.2	Les critères formels	210
10.4.3	Les noms humains collectifs	213
10.5	Les $N (E+DET:G) N:G$ et les $NI (E+DET:G) A:G N2:G$ animés non-humains...	213
10.6	Les $N (E+DET:G) N:G$ et les $NI (E+DET:G) A:G N2:G$ concrets.....	215

10.6.1	La notion sémantique « concret »	215
10.7	La classe des noms abstraits.....	218
CINQUIEME PARTIE : TRAITEMENT AUTOMATIQUE		221
Chapitre 11. L'analyse automatique des textes grecs		217
11.1	Introduction	222
11.2	Reconnaissance lexicale des $N (E+DET:G) N:G$ dans les textes	224
11.2.1	Problèmes de variantes des $N (E+DET:G) N:G$	224
11.3	Méthodes de reconnaissance lexicale automatique des $N (E+DET:G) N:G$	227
11.3.1	Les dictionnaires électroniques morphologiques	228
11.3.1.1	Structure des entrées.....	229
11.3.1.2	Génération automatique des formes fléchies	231
11.3.1.3	La flexion des unités lexicales multi-mots	232
11.3.1.3.1	Déplacement de l'accent.....	233
11.3.1.3.2	Genre des $N (E+DET:G) N:G$	234
11.3.1.3.3	Les variantes	234
11.3.1.4	Application des dictionnaires et résultats obtenus	234
11.3.2	Les règles de grammaire	236
11.3.3	Les tables du lexique-grammaire	239
11.3.3.1	Conversion automatique des tables en grammaires	247
11.3.3.2	Les graphes paramétrés	247
11.3.3.3	Application aux $N (E+DET:G) N:G$	248
Conclusion.....		253
Bibliographie.....		257

Notations

Nous nous servons ici des notations couramment utilisées dans les travaux du LADL¹ pour la description du français et des autres langues naturelles.

Les exemples cités dans la présente étude sont présentés sous trois formes, à savoir :

φακός επαφής

lentille contact-*Gfs*

(lentille de contact)

Dans la première ligne, figurent toujours les exemples grecs. Pour la commodité de la lecture par les francophones nous accompagnons systématiquement nos exemples grecs de leur traduction française. La traduction (facultative) dans la deuxième ligne traduit le contenu de l'exemple grec littéralement, tout en conservant son ordre linéaire. Les traits flexionnels des mots grecs sont systématiquement indiqués dans les traductions littérales et sont représentés par des indices alphabétiques (*i.e.* -*Nfs*, -*Ans*, etc.). Ils seront tous explicités par la suite. Les équivalences de traduction en français apparaissent dans la troisième ligne et entre parenthèses. Par principe, nous cherchons à traduire nos exemples de noms composés grecs par leurs équivalents français. Quand la traduction exacte n'existe pas, nous nous contentons de donner une traduction approximative. Enfin, lorsque les exemples grecs sont constitués de combinaisons lexicales uniquement compositionnelles (ou libres), nous donnons seulement l'équivalent de traduction en français (sans donner la traduction littérale).

Dans la présente étude, nous avons utilisé les notations suivantes :

<i>A</i>	adjectif
<i>Adj-n</i>	nom morphologiquement lié à l'adjectif
<i>AN</i>	nom composé de forme <i>Adjectif+Nom</i>
<i>C</i>	élément lexical figé (constante)
<i>Conj</i>	conjonction

¹ Laboratoire d'Automatique Documentaire et Linguistique.

<i>Conjc</i>	conjonction de coordination
<i>Ddéf</i>	article défini
<i>Ddém</i>	adjectif démonstratif
<i>Dét</i>	déterminant
<i>Dét=: E</i>	déterminant zéro (voire, absence de déterminant)
<i>Dind</i>	article indéfini
<i>Dnom</i>	déterminant nominal
<i>Dnum</i>	déterminant numéral (adjectival ou nominal)
<i>E</i>	séquence vide
<i>GN</i>	groupe nominal
<i>N (E+DET:G) N:G</i>	nom composé de structure <i>N (E+Déterminant au génitif) Nom au génitif</i>
<i>N Prép N</i>	nom composé de structure <i>Nom+Préposition+Nom</i>
<i>N</i>	substantif
<i>N+A</i>	adjectif substantivé
<i>N₀</i>	sujet
<i>N₁</i>	premier complément variable
<i>N₂</i>	second complément variable
<i>Nabs</i>	nom de type ‘abstrait’
<i>Nanim</i>	nom de type ‘animé’
<i>Nconc</i>	nom de type ‘concret’
<i>Ncomp</i>	nom composé
<i>Nhum</i>	nom de type ‘humain’
<i>Nhumcoll</i>	nom de type ‘humain collectif’
<i>NN</i>	nom composé de structure <i>Nom+Nom</i>
<i>Npréd</i>	nom prédicatif
<i>P</i>	phrase élémentaire
<i>Poss_c</i>	adjectif possessif composé
<i>Poss_s</i>	adjectif possessif simple
<i>Possⁱ</i>	adjectif possessif (catégorie grammaticale)
	(l’exposant indique la coréférence obligatoire à <i>N_i</i>)
<i>Prép</i>	préposition
<i>Pro</i>	pronom

<i>V</i>	verbe
<i>Vsup</i>	verbe support
<i>W</i>	séquence de compléments quelconques

Les informations flexionnelles apparaissent en indices à droite de *N* ou de *A*. Elles sont constituées de trois lettres : la première signale le cas, la deuxième le genre et la dernière le nombre. Plus explicitement, les notations concernant les cas sont :

<i>N</i>	nominatif
<i>G</i>	génitif
<i>A</i>	accusatif
<i>V</i>	vocatif

Les notations désignant les genres sont :

<i>m</i>	masculin
<i>f</i>	féminin
<i>n</i>	neutre

Enfin, les notations indiquant les nombres sont :

<i>s</i>	singulier
<i>p</i>	pluriel

Par exemple, l'indice *-Nfs* signifie que la forme nominale est au nominatif du féminin, du singulier. L'indice *-Anp* signifie que la forme nominale est à l'accusatif du neutre, du pluriel, et ainsi de suite.

Enfin, nous nous servons des signes conventionnels suivants :

*	suite de mots ou phrase non acceptables, ou éventuellement acceptables dans un emploi différent de celui envisagé
?	suite de mots ou phrase d'acceptabilité douteuse
?*	suite de mots ou phrase proche de l'inacceptabilité
=	équivalence sémantique
≠	absence d'équivalence sémantique entre deux formes

=:	explicitation lexicale ou syntaxique ou structurelle du contenu d'une forme donnée
≠:	explicitation du contenu d'une forme donnée en une forme lexicale ou syntaxique ou structurelle différente de celle indiquée à gauche du symbole
#	changement de phrase (découpage de phrase complexe)
($x+y$)	possibilité de choix entre les éléments x et y de la parenthèse
[x z .]	réduction de l'élément x

Introduction

L'objectif de cette étude est l'élaboration d'un dictionnaire électronique des noms composés grecs de structure *Nom Nom-Génitif* (e.g. ζώνη ασφαλείας/ceinture de sécurité) en vue de leur reconnaissance lexicale automatique dans les textes écrits. Cependant, les ressources proposées pourraient être utilisées aussi dans d'autres applications en traitement automatique des langues (TAL), à savoir : analyse automatique des textes, recherche et extraction d'information, indexation des textes, applications multilingues, etc.

Dans le cadre de cette recherche, nous étudions les noms composés du grec moderne qui sont définis par la structure morphosyntaxique interne : *Nom (E + Déterminant-Génitif) Nom-Génitif*. (désormais notés : *N (E + DET:G) N:G*). En gros, ils correspondent aux noms composés du français de structure : *Nom de Nom* (A. Firenze et B. Pelletier 1989, G. Gross 1991, G. Gross et A. Dugas 1991, A. Belanger 1997). A titre d'illustration, citons les exemples suivants :

NN:G

φακός επαφής

lentille contact-*Gfs*

(lentille de contact)

N (E + DET:G) N:G

πιάτο (E + της) ημέρας

plat (E + la-*Gfs*) journée-*Gfs*

(plat du jour)

N DET:G N:G

κουταλιά της σούπας

cuillerée la-*Gfs* soupe-*Gfs*

(cuillerée à soupe)

NN:G (= A N A:G N:G)²

νομικό πρόσωπο δημοσίου δικαίου

juridique personne public-*Gns* droit-*Gns*

(personne morale de droit public)

² Il s'agit d'un nom composé complexe (ou surcomposé) de structure *NN:G* dont les deux composants sont eux-aussi des noms composés binaires de structure *AN*. Dans le cadre de ce travail, nous étudions également les noms composés complexes (ou surcomposés) (cf. II, 5.3 et III, 6.3).

Comme le note M. Silberztein (1993 : 405), parmi les unités de la langue, les noms composés sont les plus nombreux, étant donné que le vocabulaire de base contient cinq fois plus de noms composés que de noms simples, toutes parties du discours confondues. Il est donc indispensable de recenser et décrire ces unités souvent mal répertoriées dans les dictionnaires usuels.

Notre méthodologie s'appuie sur la recherche de critères formels d'ordre morphologique, syntaxique et sémantique afin de définir et recenser les noms composés du grec moderne tout en les distinguant des syntagmes nominaux libres de même structure. Mis à part les noms composés entièrement figés ou lexicalisés³, qui sont les moins nombreux dans l'ensemble des candidats noms composés recensés, il est souvent délicat de décider sur le statut d'une séquence repérée.

Nous délimitons les séquences polylexicales rassemblées aux catégories suivantes :

- les noms composés entièrement figés (e.g. *έξοδος κινδύνου*/issue de secours, *φακός επαφής*/lentille de contact) ;
- les syntagmes nominaux semi-figés (ou productifs) (e.g. *κέντρο (αισθητικής + θαλασσοθεραπείας + ερευνών + τηλεπικοινωνιών + αποτοξίνωσης + ...)*/centre de (beauté + thalassothérapie + recherche + télécommunications + désintoxication + ...)).
- les entités nommées (cf. N. Chinchor 1998), (e.g. *Παγκόσμια Οργάνωση Υγείας*/Organisation Mondiale de la Santé, *Ευρωπαϊκή Κοινότητα Άνθρακα και Χάλυβα*/Communauté européenne du charbon et de l'acier) ;
- les termes de spécialité (cf. T. Cabré 1998), (e.g. *επίσχεση ούρων*/rétention d'urine (domaine : Médecine), *νόσφιση εξουσίας*/usurpation de pouvoir (domaine : Droit), *σήμα ομίχλης*/signal de brume (domaine : Transport maritime) ;
- et les syntagmes nominaux figés qui font partie d'une phrase simple figée et qui ont un emploi autonome en dehors de l'expression figée. Par exemple, le *N (E+DET:G) N:G=*: *βάπτισμα του πυρός*/baptême du feu fait partie de l'expression figée :

³ Les noms entièrement figés sont lexicalisés. Nous utilisons indifféremment les termes « nom composé entièrement figé » et « nom composé lexicalisé ».

Η Μαρία πήρε το βάπτισμα του πυρός

Maria a reçu le baptême du feu

mais il peut avoir également un emploi autonome en dehors de la phrase figée, étant donné qu'il peut apparaître aussi dans des phrases compositionnelles comme :

Το βάπτισμα του πυρός αργεί

Le baptême du feu tarde

Nous distinguons ces $N(E+DET:G)N:G$ des syntagmes nominaux figés qui font partie d'une phrase simple figée et qui n'ont pas d'emploi autonome en dehors de l'expression figée (cf. M. Gross 1982, 1989), (e.g. *κρούω τον κώδωνα του κινδύνου*/sonner le tocsin) :

Η Ιταλία κρούει τον κώδωνα του κινδύνου για την άνοδο της θερμοκρασίας

L'Italie sonne l'alarme le-*Gms* danger-*Gms* pour la montée la température

(L'Italie sonne le tocsin pour la montée de la température)

* *Ο κώδων του κινδύνου που κρούει η Ιταλία για την άνοδο της θερμοκρασίας*

* L'alarme le-*Gms* danger-*Gms* que sonne l'Italie pour la montée la température

(* Le tocsin que sonne l'Italie pour la montée de la température)

Ces derniers ne font pas l'objet de la présente étude.

Notre choix d'étudier les noms composés de structure $N(E + DET:G)N:G$ du grec moderne se justifie par trois raisons :

- leur nombre élevé : pour ce qui est du grec moderne, les $N(E + DET:G)N:G$ constituent avec les noms composés de structure AN les deux catégories les plus chiffrées (cf. T. Kyriacopoulou ; S. Mrabti et A. Yannacopoulou 2002⁴) ;
- l'existence de quelques critères formels : il existe des critères plus ou moins précis qui permettent de les recenser et de les étudier ;
- le manque d'étude systématique : les noms composés de structure $N(E + DET:G)N:G$ du grec moderne n'ont jamais été étudiés intégralement et systématiquement. Cependant, des études de cas existent et sont discutées dans I, 1.1.

⁴ Désormais T. Kyriacopoulou *et al.* (2002).

A ces trois raisons s'ajoute le fait que les noms composés font partie des expressions « multi-mots » (en anglais, *multiword expressions*), qui se trouvent actuellement au centre des préoccupations de la communauté scientifique internationale. Elles ont déjà fait l'objet, à plusieurs reprises, de conférences et de groupes de travail internationaux, comme par exemple : les conférences ACL (2003⁵, 2004⁶, 2007⁷), EACL (2006⁸), LREC (2008⁹), COLING/ACL (2006¹⁰).

Notre travail s'inscrit dans la perspective de l'élaboration du DELAC et du lexique-grammaire général du grec moderne en vue de l'analyse automatique des textes écrits. Les noms composés posent problème lors de l'analyse lexicale des textes, qui constitue l'étape préalable à toute analyse ou génération ou encore traduction automatiques des textes. En d'autres termes, si un nom composé n'est pas reconnu en tant qu'une seule unité lexicale, il ne pourra pas être analysé correctement et, par conséquent, ne sera pas proprement traduit. La solution à ce problème, qui semble être la plus adéquate actuellement, consiste à recenser les noms composés de manière exhaustive et systématique sur la base de critères formels (M. Silberztein 1990, C. Mota ; P. Carvalho et E. Ranchhod 2004, E. Ranchhod 2005). Une fois les noms composés recensés, il est nécessaire d'étudier et représenter formellement toutes les informations morphologiques, syntaxiques et sémantiques les concernant, afin qu'ils puissent être analysés automatiquement.

Cependant, une telle approche engendre des problèmes qui touchent, d'une part, à la définition et délimitation des noms composés ainsi qu'aux méthodes de leur représentation formalisée et, d'autre part, à la maintenance et la mise à jour des ressources construites. Dans le cadre de la présente étude, l'objectif que nous nous sommes fixée consiste à étudier les propriétés spécifiques des différents $N(E + DET:G) N:G$ recensés afin d'en repérer des sous-catégories opérationnelles en vue de l'analyse automatique des textes. Dans ce but, le recensement, la description linguistique et la représentation formalisée des données en vue de leur reconnaissance lexicale automatique sont effectués en fonction de cette répartition. Ceci permet, tout d'abord, de mettre de l'ordre dans le dictionnaire électronique des noms composés qui existe actuellement pour le grec moderne (cf. T. Kyriacopoulou *et al.* 2002) et

⁵ ACL workshop *Multiword Expressions: Analysis, Acquisition and Treatment*.

⁶ ACL workshop *Multiword Expressions: Integrating Processing*.

⁷ ACL workshop *A Broader Perspective on Multiword Expressions*.

⁸ EACL workshop *Multi-word-expressions in a multilingual context*.

⁹ LREC workshop *Towards a Shared Task for Multiword Expressions*.

¹⁰ ACL workshop *Multiword Expressions: Identifying and Exploiting Underlying Properties*.

qui constitue aujourd'hui l'unique méthode de représentation formalisée adoptée pour l'ensemble des noms « mutli-mots ». De plus, ceci nous permet de faire une proposition pour l'organisation et la structuration des données qui servira de base pour leur enrichissement continu et leur exploitation en vue des différentes applications du TAL.

L'équipe des chercheurs qui travaillent sur le grec moderne a entrepris une description formalisée à grande échelle de la langue grecque, comme ceci a été fait pour le français et pour d'autres langues (italien, espagnol, portugais, allemand, etc.). Cette recherche a conduit à la construction d'un système de dictionnaires électroniques, dans lesquels les unités de la langue sont recensées et leurs propriétés morphologiques, syntaxiques et sémantiques sont décrites de manière exhaustive et systématique. Du point de vue linguistique, les principes théoriques et méthodologiques, qui régissent notre entreprise, s'appuient sur la grammaire transformationnelle harisienne (cf. Z. S. Harris 1964, 1968, 1976), dans la version qu'en a proposé M. Gross (1975, 1977). Ces principes, étant exhaustivement explicités chez M. Gross (1988a), J.-P. Boons ; A. Guillet et C. Leclère (1976) et les autres chercheurs¹¹ du LADL, il n'est pas nécessaire d'y revenir ici en détail. Du point de vue applicatif, nous allons présenter les problèmes et les solutions éventuellement proposées afin de pouvoir utiliser ces données dans l'analyse lexicale des textes écrits.

Cette étude s'organise en cinq parties.

Dans la première partie, nous délimitons l'objet de notre étude. Après avoir présenté un survol de la littérature grammaticale sur la composition nominale en grec moderne et en français, nous essayons de définir la notion de figement qui régit cette étude et nous exposons les critères morphologiques, syntaxiques et sémantiques que nous utilisons dans ce but.

La deuxième partie est consacrée à la présentation de nos méthodes de recensement et à l'étude de la variation au sein des $N(E + DET:G)N:G$. Des problèmes particuliers, concernant la représentation formalisée des noms composés dans le dictionnaire électronique morphologique et dans les tables du lexique-grammaire, ainsi que les solutions éventuellement adoptées, sont également discutés dans cette partie.

¹¹ Cf. C. Leclère et C. Subirats-Rüggeberg (1991).

L'étude de la structure lexicale interne des $N(E + DET:G)N:G$ constitue l'objet de la troisième partie. Nous présentons ici les différentes sous-catégories des $N(E + DET:G)N:G$ identifiées lors de l'étape du recensement et nous étudions la structure de leurs constituants.

La quatrième partie porte sur l'étude linguistique des $N(E + DET:G)N:G$. Nous effectuons une étude syntaxico-sémantique du génitif au sein des $N(E + DET:G)N:G$ et la distinction entre $N(E + DET:G)N:G$ prédicatifs et non-prédicatifs. Nous abordons également les propriétés sémantiques qui sont codées systématiquement dans les tables du lexique-grammaire.

La cinquième partie traite l'exploitation de nos données linguistiques en vue de leur reconnaissance lexicale automatique. Nous présentons ici les différentes méthodes de représentation formalisée que nous avons choisies pour nos données.

Enfin, nous faisons figurer en annexe (cf. Annexes, Volume 2) des échantillons représentatifs du dictionnaire électronique morphologique des noms composés, des tables du lexique-grammaire, des grammaires locales de reconnaissance des syntagmes nominaux semi-figés et de la liste d'entités nommées.

PREMIERE PARTIE : DELIMITATION DE L'ETUDE

Chapitre 1. Travaux antérieurs

1.1 Survol de la littérature sur les noms composés en grec moderne

Un nombre important de travaux ont été consacrés à la composition lexicale (ou morphologique) en général, et plus particulièrement à la composition nominale du grec moderne. Dans ce sous-chapitre, nous citons les travaux les plus significatifs qui ont été effectués dans le cadre de l'approche linguistique et qui servent aujourd'hui de référence pour la plupart des études portant sur la composition nominale en grec moderne.

Plusieurs auteurs ont déjà proposé des survols des travaux sur les noms composés en grec moderne notamment dans le cadre de l'approche linguistique (ou symbolique) (cf. Z. Gavriilidou 1997, A. Anastassiadis-Syméonidis 2004). Dans cette section, nous ne reprenons que les principales étapes de la réflexion sur la composition en grec moderne. Nous décrivons les principes qui régissent la composition nominale du grec moderne et nous présentons une typologie de classification.

Il faut rappeler que la langue grecque dans son état actuel comprend le grec populaire (ou la *démotique*¹²) ainsi que des résidus du grec ancien et du grec savant (ou *katharevoussa*¹³), ce que G. Gross (1996a : 22) appelle des « blocs erratiques », c'est-à-dire « des éléments ou constructions qui remontent à un état de langue antérieur ». La somme des trois constitue ce qu'on appelle largement *le grec moderne*. Malgré le fait que la *démotique* est depuis plus de 30 ans la langue officielle en Grèce, des résidus de la langue savante et du grec ancien persistent, ce qui rend le grec une langue riche en variantes et en mots composés.

Contrairement à la grande diversité des critères de définition concernant la composition nominale en français (cf. I, 1.2), en grec moderne, les noms composés sont essentiellement définis sur la base de critères morphologiques (Z. Gavriilidou 1997). Cependant, les linguistes et les grammairiens grecs se servent également d'un certain nombre de critères syntaxiques et sémantiques afin de définir les noms composés non-soudés, qui sont désignés sous différentes appellations, à savoir : noms *phrastiques*¹⁴ (Ch. Clairis et G. Babiniotis 1996), combinaisons

¹² La *démotique* est la forme standard du grec moderne. Elle est devenue la langue officielle de la Grèce en 1976.

¹³ Langue officielle de la Grèce jusqu'en 1976.

¹⁴ Nous traduisons le terme grec *φραστικά ονόματα* par le terme français *noms phrastiques*.

figées¹⁵ ou noms composés juxtaposés lâches¹⁶ (D. Holton ; P. Mackridge et I. Philippaki-Warburton 2000¹⁷), noms composés lâches¹⁸ (A. Tsopanakis 1994), noms composés polylexicaux lâches¹⁹ (A. Ralli 2004).

M. Triantaphyllidis (1941), dans sa grammaire intitulée *Grammaire néohellénique de la démotique*, consacre une grande partie aux noms composés soudés. Tout d’abord, il distingue les noms dérivés des noms composés : les noms dérivés sont formés à partir d’un suffixe ajouté à la base lexicale (mot racine) (e.g. τραγουδώ-τραγουδιστής/chanter-chanteur, ιδρύω-ιδρυμα/fonder-fondation, γάλα-γαλατάς/lait-laitier) d’un mot simple, tandis que les noms composés sont formés à partir de l’union des bases lexicales (mots racines) de deux – ou plusieurs – mots simples ou dérivés (e.g. ανεμόμυλος/moulin à vent, αγριολούλουδο/fleur des champs, πανωφόρι/manteau). Le même auteur décrit explicitement les phénomènes de dérivation lexicale (ou morphologique) et de composition lexicale en grec moderne tout en mettant l’accent sur les principes qui les régissent. Plus particulièrement, concernant la dérivation nominale, il remarque que les noms peuvent dériver de verbes, de noms et d’adjectifs. Il donne la liste quasi-exhaustive des différents suffixes mis en jeu et de la valeur sémantique qu’ils apportent à l’unité lexicale dérivée. Quant à la composition nominale, il propose tout d’abord une classification des mots composés sur la base de leur sens : il les distingue donc en juxtaposés²⁰, dont le sens est la somme de leurs deux composants (e.g. ρυζόγαλο/riz au lait-ρύζι/riz και/et γάλα/lait) ; déterminatifs²¹, dont le premier composant devient complément (nominal, adjectival ou adverbial) du deuxième lors de leur analyse (e.g. κρασοπότηρο-ποτήρι του κρασιού/verre à vin) ; possessifs²², qui désignent celui qui possède quelque chose (e.g. καλότυχος-‘εκείνος που έχει καλή τύχη’/chanceux-‘celui qui a de la chance’) ; et objectifs²³, dont l’un composant constitue le complément d’objet direct de l’autre (e.g. χαρτοπαίχτης/joueur aux cartes). Ensuite, il décrit le comportement morphologique des mots composés en insistant, plus spécifiquement, sur les phénomènes d’accentuation, de

¹⁵ Nous traduisons le terme grec στερεότυποι συνδυασμοί par le terme français *combinaisons figées*.

¹⁶ Nous traduisons le terme grec χαλαρά παραθετικά σύνθετα par le terme français *noms composés juxtaposés lâches*.

¹⁷ Désormais D. Holton *et al.* (2000).

¹⁸ Nous traduisons le terme grec χαλαρά σύνθετα par le terme français *noms composés lâches*.

¹⁹ Nous traduisons le terme grec χαλαρά πολυλεκτικά σύνθετα par le terme français *noms composés polylexicaux lâches*.

²⁰ Nous traduisons le terme grec παρατακτικά par le terme français *juxtaposés*.

²¹ Nous traduisons le terme grec προσδιοριστικά par le terme français *déterminatifs*.

²² Nous traduisons le terme grec κτητικά par le terme français *possessifs*.

²³ Nous traduisons le terme grec αντικειμενικά par le terme français *objectifs*.

composition nominale savante et populaire et de sur-composition (*i.e.* formation des mots composés à plus de deux composants).

Dans la grammaire de M. Triantaphyllidis (1941), les noms composés non-soudés ne sont pas abordés. L'auteur (2005 : 178) signale que la composition (soudée) n'est pas un phénomène qui a lieu d'un seul coup. Le sentiment qu'il s'agisse d'un nom composé soudé peut tout simplement être expliqué par le fait qu'il existe des paires de mots fréquemment juxtaposés dans la phrase. Les limites entre la juxtaposition et la composition ne sont donc pas toujours claires. Pour cette raison, on retrouve souvent les mêmes combinaisons de mots qui sont écrites parfois en un seul mot (e.g. *καληνύχτα*/bonne nuit), parfois en deux (ou plusieurs) mots (e.g. *καλή νύχτα*/bonne nuit) ou, encore, des dérivés de ces combinaisons (e.g. *καληνυχτίζω*/faire la bonne nuit).

Selon A. Anastassiadis-Syméonidis (1986 : 138-143), qui effectue la première étude approfondie sur les noms composés polylexicaux, il existe six types de critères, qui permettent de constater si une séquence donnée constitue une séquence libre ou bien un nom composé : 1) phonologiques, 2) graphiques, 3) morphologiques, 4) syntaxiques, 5) sémantiques et 6) d'emploi.

Phonologiquement, les noms composés constituent une seule unité au niveau de la respiration, avec deux ou plusieurs accents dynamiques. Au niveau graphique, les composants ne sont pas soudés. Au niveau morphologique, les constituants d'un nom composé préservent leur forme complète au sein du nom composé. A. Anastassiadis-Syméonidis (1986 : 139) précise que les constituants des noms composés ne recouvrent pas toutes les catégories grammaticales, à savoir les conjonctions, les interjections, les verbes auxiliaires, les pronoms et l'article indéfini n'y sont pas présents. Elle ajoute qu'au niveau morphosyntaxique le nom composé est une séquence de mots linéaire, mais qui cache une structure hiérarchique.

Quant au niveau syntaxique, l'ordre des constituants est fixe au sein du nom composé :

<i>κοινωνική λειτουργός</i>	<i>*λειτουργός κοινωνική</i>
assistant social	*social assistant
<i>φακοί επαφής</i>	<i>*επαφής φακοί</i>
lentilles de contact	*de contact lentilles

Il est impossible de remplacer un seul élément du nom composé par un pronom, puisque les composants perdent leur autonomie au sein du nom composé. Enfin, A. Anastassiadis-Syméonidis (1986 : 140) signale la création des dérivés dont la base est le nom composé dans son ensemble :

A	N	AN-a
	ἀκρα αριστερά	→ ακροαριστερός, -ή, -ό
	extrême gauche	→ extrême gauchiste

Au niveau sémantique, il est possible de remplacer un constituant du nom composé ou bien le nom composé dans son ensemble par un autre mot :

μικρή οθόνη	= τηλεόραση
petit écran	= télévision

A. Anastassiadis-Syméonidis (1986 : 141) souligne que l'impossibilité de prévoir le sens du nom composé à partir du sens de ses composants dépend du niveau de lexicalisation du nom composé.

Enfin, concernant le critère de l'emploi, A. Anastassiadis-Syméonidis (1986 : 142) note que la notion de l'emploi est étroitement liée à la fréquence dans l'emploi, tout en précisant cependant que celui-ci ne s'avère pas un critère efficace concernant la distinction entre noms composés et séquences libres de même structure.

Dans *Grammaire grecque moderne*, A. Tsopanakis (1994) n'aborde que très brièvement le phénomène de la composition nominale non-soudée, pour laquelle il propose la dénomination *composition lâche*²⁴. Il remarque que la composition lâche est l'emploi fréquent et figé de deux mots simples, le plus souvent d'un adjectif et d'un nom, comme par exemple *Άγιον Όρος*/Mont Athos, *Βόρεια Ελλάδα*/Grèce du Nord, *Βόρειος Πόλος*/Pôle Nord, *Μέγας Αλέξανδρος*/Alexandre le Grand etc. Il note que les mots simples qui constituent les noms composés lâches s'écrivent séparément, mais leur combinatoire restreinte fait qu'ils forment ensemble un sens unique. Ceci est démontré également par la formation de noms dérivés soudés provenant de ces noms composés lâches, par exemple *αγιορείτης*/habitant du Mont Athos, *βορειοελλαδίτης*/habitant de la Grèce du Nord.

²⁴ Nous traduisons le terme grec *χαλαρή σύνθεση* par le terme français *composition lâche*.

Notons que tous les exemples des noms composés lâches de structure adjectif + nom cités par A. Tsopanakis (1994) sont des entités nommées (T. Poibeu 2005, I. Michailidis *et al.* 2006), c'est-à-dire des noms propres de personnes, de lieux ou d'organisations. Il ne donne pas d'exemples de noms composés communs de la même structure. De plus, il décrit uniquement les noms composés de structure adjectif + nom et ne fait aucune référence aux noms composés de structure différente, par exemple nom + nom, nom + nom au génitif etc.

A. Anastassiadis-Syméonidis (1994) présente des cas d'asymétrie lexicale et structurelle entre les noms composés de structure *Adjectif + Nom* (e.g. παιδική χαρά/aire de jeux) en français et en grec, ainsi que les indices de figement permettant de distinguer les noms composés *Adjectif + Nom* des séquences compositionnelles de même structure (A. Anastassiadis-Syméonidis 2004 : 12-13).

A. Anastassiadis-Syméonidis (1996 : 98-100) raffine les critères syntaxiques et sémantiques concernant la distinction entre noms composés polylexicaux et séquences libres de même structure.

Ch. Clairis et G. Babinotis (1996) abordent plus explicitement le phénomène de la composition en grec moderne dans leur ouvrage *Grammaire du grec moderne*. Ils ne s'intéressent pas uniquement à l'étude des critères morphologiques, mais ils explorent également un certain nombre de critères syntaxiques et sémantiques qui jouent un rôle important dans la définition de la composition.

En ce qui concerne les noms composés soudés, les auteurs considèrent que les noms composés du grec moderne résultent de la combinaison de mots simples appartenant aux catégories grammaticales suivantes :

nom + nom

άνεμος + μύλος → ανεμόμυλος

vent + moulin → moulin à vent

adjectif + nom

άγριος + άνθρωπος → αγριάνθρωπος

sauvage + homme → homme sauvage (renfermé)

adverbe + nom

πάνω + σεντόνι → πανωσέντονο

au dessus + drap → drap du dessus

pronom + nom

αυτός + κάθαρση → αυτοκάθαρση

soi + purification → autopurification

déterminant numérique + nom

δέκα + δραχμή → δεκάδραχμο

dix + drachmes → pièce de dix drachmes

Selon Ch. Clairis et G. Babiniotis (1996), les noms composés se divisent, en fonction de la relation syntaxique et sémantique qui régit leurs différents composants, en deux grandes catégories : les noms composés juxtaposés²⁵ et les noms composés dépendants²⁶. Le sens des noms composés juxtaposés devient évident si on place la conjonction de coordination *και*/et entre les différents composants obligatoirement nominaux. La possibilité de coordonner les composants montre que ces derniers sont syntaxiquement et sémantiquement autonomes. Par exemple :

αντρώγυννο → άντρας και γυναίκα

(couple marié) → homme et femme

παρασκευοσαββατοκύριακο → Παρασκευή και Σάββατο και Κυριακή

(long week-end) → vendredi et samedi et dimanche

En revanche, les noms composés dépendants expriment une relation de dépendance entre leurs différents composants (e.g. *αγριόπαπια*/canard sauvage, *ηλιοβασίλεμα*/coucher de soleil). Le nom-tête (ou noyau) est par principe postposé ; le premier composant est complément du nom-tête (e.g. *η αγριότητα μιας πάπιας*/la sauvagerie d'un canard, *το βασίλεμα του ήλιου*/le coucher du soleil). En fonction de ces relations syntaxiques et sémantiques, les noms composés dépendants se distinguent, à leur tour, en noms composés qualificatifs²⁷ (e.g. *μια πάπια που είναι άγρια*/un canard qui est sauvage) et en noms composés verbaux²⁸ (e.g. *ο ήλιος βασιλεύει*/le soleil se couche). Les auteurs (1996 : 98) citent uniquement des exemples de noms composés soudés.

Chez Ch. Clairis et G. Babiniotis (1996), les noms composés non-soudés sont désignés sous le terme de « noms phrastiques ». Ils signalent que les « noms phrastiques » sont courants en grec moderne et qu'ils présentent les structures suivantes :

²⁵ Nous traduisons le terme grec *παρατακτικά σύνθετα* par le terme français *noms composés juxtaposés*.

²⁶ Nous traduisons le terme grec *εξαρτημένα σύνθετα* par le terme français *noms composés dépendants*.

²⁷ Nous traduisons le terme grec *προσδιοριστικά σύνθετα* par le terme français *noms composés qualificatifs*.

²⁸ Nous traduisons le terme grec *ρηματικά σύνθετα* par le terme français *noms composés verbaux*.

- nom + nom au génitif (φακοί επαφής/lentilles de contact, τιμή ασφαλείας/prix de retient) ;
- adjectif + nom (παιδική χαρά/aire de jeux, πυρηνικός πόλεμος/guerre nucléaire) ;
- nom + nom (παιδί-θαύμα/enfant prodige, λέξη-κλειδί/mot-clé).

Ils notent que les « noms phrastiques » sont caractérisés par une autonomie morphologique et phonologique, mais pas par une autonomie syntaxique, et par conséquent, les parties qui les constituent ne peuvent pas être actualisées (déterminées ou modifiées) individuellement, mais uniquement en tant qu'ensemble :

παιδική χαρά	*παιδική μεγάλη χαρά	mais : μεγάλη παιδική χαρά
aire de jeux	*aire de grands jeux	mais : grande aire de jeux

Lorsque la coordination des modifieurs du même type est autorisée, on a affaire à un nom composé comportant deux (ou plusieurs) modifieurs :

κρέμα ενυδάτωσης και αντηλιακής προστασίας
 crème hydratation._{Gfs} et contre-soleil._{Gfs} protection._{Gfs}
 (crème hydratante et protection soleil)

Dans le cas où cette coordination n'est pas autorisée, le modifieur inséré ne fait pas partie du nom composé :

παιδική χαρά	*μεγάλη και παιδική χαρά
aire de jeux	*aire de jeux et de récréation

La permutation n'est souvent pas permise dans le cas des « noms phrastiques » :

παιδική χαρά	*χαρά παιδική
aire de jeux	*de jeux aire

Dans certains « noms phrastiques » de structure nom + nom au génitif, le deuxième composant nominal peut être déterminé par l'article défini :

κουτάλακι του γλυκού
 cuillère le-_{Gs} dessert-_{Gs}
 (cuillère à dessert)

Lorsque le premier composant nominal des « noms phrastiques » de structure nom + nom au génitif est omis, le deuxième composant nominal au génitif manifeste un emploi prédicatif. Par exemple :

είναι (Ε + προϊόν) εισαγωγής/est (un produit) d'importation

είναι (Ε + ρούχο) της μόδας/est (un vêtement) à la mode

Certains « noms phrastiques » de structure adjectif + nom s'emploient uniquement au pluriel :

καλές τέχνες/beaux arts

**καλή τέχνη/bel art*

D'autres donnent lieu aussi à des dérivés soudés, en l'occurrence des adjectifs :

<i>A</i>	<i>N</i>	<i>AN-a</i>
<i>Αιγαίο Πέλαγος</i>	\rightarrow	<i>αιγαιοπελαγίτικος, -η, -ο</i>
Mer Egée	\rightarrow	de la Mer Egée

Enfin, Ch. Clairis et G. Babiniotis (1996) consacrent un paragraphe aux sigles. Selon eux, les sigles sont une sous-catégorie des « noms phrastiques » et constituent un type d'unité lexicale abrégée, étant donné qu'ils sont formés à partir :

- des lettres initiales des mots simples capitalisés d'un syntagme nominal (en excluant les articles et les conjonctions) :

O.T.E. (Οργανισμός Τηλεπικοινωνιών Ελλάδος)
O.T.E. (Organisation de Télécommunications de Grèce)

- des ensembles²⁹ de lettres initiales (combinaisons des consonnes et voyelles) :

ΕΛ.ΚΕ.ΠΑ. (Ελληνικό Κέντρο Παραγωγικότητας)
EL.KE.PA. (Centre Grec de Productivité)

- d'une combinaison des lettres initiales et des ensembles (combinaisons des consonnes et voyelles) :

²⁹ Nous traduisons le terme grec *σύμπλεγμα* par le terme français *ensemble*.

Ο.ΠΕ.Κ. (Οργανισμός Πετρελαιοπαραγωγικών Κρατών)

OPEP (Organisation des pays exportateurs de pétrole)

Ils précisent que les sigles peuvent effectivement être vus comme des abréviations comportant des points en tant que séparateurs de leurs composants. Ils sont des produits de la langue écrite, mais ceux qui sont très fréquents sont également utilisés au discours oral. Du point de vue morphologique, les sigles sont invariables³⁰. Dans la majorité des cas, leur genre est déterminé par le genre du nom-tête du sigle³¹ :

ο Ο.Τ.Ε. (Οργανισμός Τηλεπικοινωνιών Ελλάδος), όπου η βασική λέξη είναι Οργανισμός

O.T.E. (**masculin**) (Organisation de Télécommunications de Grèce), où le nom-tête du sigle est **Organisation** (en grec, genre : masculin)

Enfin, ils remarquent que les sigles très usités peuvent donner lieu à des mots dérivés, comme par exemple :

οτετζής/employé de l'O.T.E.

Dans *Grammaire de la langue grecque*, D. Holton *et al.* (2000) ne se réfèrent que très brièvement à la formation de mots composés en grec moderne. Ils remarquent que la langue grecque dispose d'un grand nombre de mots composés (soudés), qui sont formés à partir des bases lexicales (mots racines) de deux ou plusieurs mots simples distincts (D. Holton *et al.* 2000 : 184). Plus particulièrement, pour ce qui est de la composition nominale, ils précisent qu'il y a plusieurs types de noms composés : nom + nom, nom + verbe, adjectif + nom etc. (D. Holton *et al.* 2000 : 185). La soudure des différents composants des noms composés s'effectue le plus souvent au moyen de la voyelle *-o-*, comme dans les exemples :

$$N + V \quad N_{comp}$$

βίος + γράφω → βιογραφία
(vie + écrire → biographie)

³⁰ Ce n'est pas toujours le cas. Sur les sigles qui ne sont pas invariables, cf. A. Kalampokis et T. Kyriacopoulou 2005.

³¹ On observe également des cas où le genre du sigle se différencie du genre du nom-tête du sigle. Par exemple : *(ο+το) Φ.Π.Α. (Φόρος Προστιθέμενης Αξίας), όπου η βασική λέξη είναι Φόρος* TVA (**masculin ou neutre**), (taxe à la valeur ajoutée), où le nom tête du sigle est **Taxe** (**en grec, genre : masculin**) (cf. A. Mitska et I. Michailidis 2003)

$N + N \quad N_{comp}$
 σκηνή + θέση → σκηνοθεσία
 (scène + mise → mise en scène)

En ce qui concerne les noms composés non-soudés, D. Holton *et al.* (2000 : 335) se limitent à la description des noms composés de structure nom + nom, qui constituent, selon eux, des « combinaisons figées » ou des « noms composés juxtaposés lâches ». Ils ne font aucune allusion aux noms composés non-soudés de structure différente, par exemple adjectif + nom ou nom + nom au génitif, malgré le nombre élevé de ces derniers dans le vocabulaire de la langue grecque.

Plus précisément, selon D. Holton *et al.* (2000 : 335), il existe un petit nombre de constructions figées en grec moderne qui se combinent avec le prédicat *είμαι*/être (et ses variantes lexicales et aspectuelles), mais qui peuvent également avoir un emploi exclamatif quand elles sont utilisées toutes seules (e.g. *κίνδυνος θάνατος*/danger de mort) :

Η χοληστερίνη είναι κίνδυνος θάνατος για την υγεία
 La cholestérol-*Nfs* est danger-*Nms* mort-*Nms* pour la santé
 (Le cholestérol est **un danger de mort** pour la santé)

Κίνδυνος θάνατος!
 Danger-*Nms* mort-*Nms* !
 (Danger de mort !)

D. Holton *et al.* (2000 : 335) signalent qu'il existe également un nombre croissant de « noms composés juxtaposés lâches », dont certains sont des calques d'autres langues européennes (e.g. *κράτος-μέλος*/pays-membre en français, « member state » en anglais). Alors que les expressions « traditionnelles » citées plus haut (e.g. *κίνδυνος θάνατος*/danger de mort) s'écrivent en deux mots simples séparés sans trait d'union, les combinaisons plus récentes s'écrivent avec un trait d'union (e.g. *ταξίδι-αστραπή*/voyage-éclair).

A. Ralli (2004), quant à elle, définit la composition comme la procédure qui crée des éléments morphologiquement complexes à partir d'au moins deux bases lexicales (mots racines). La composition est une procédure très productive en grec moderne, utilisée notamment dans la production des néologismes. Les mots composés ont toutes les caractéristiques de l'élément

que l'on appelle « mot » (A. Ralli 2004 : 164). Quand elle parle de « composition » et de « mots composés » elle se réfère essentiellement aux mots composés soudés.

Elle décrit les caractéristiques générales des mots composés soudés, tels que l'accent unique, le sens particulier du mot composé qui ne peut pas toujours être déduit du sens ordinaire de ses composants, ainsi que le fait qu'ils contiennent des composants qui ne sont pas toujours des mots complets. Elle parle également de la présence de la voyelle -o- (cf. D. Holton *et al.* 2000), qui, étant une caractéristique de base, constitue un indice fort de compositionnalité.

De plus, elle propose une sous-catégorisation morphosyntaxique des noms composés, qui s'aligne avec celle proposée par Ch. Clairis et G. Babinotis (1996). Elle étudie aussi la structure et l'accentuation des mots composés soudés ainsi que les noms composés qui contiennent des éléments savants.

Les noms composés non-soudés sont également abordés chez A. Ralli (2004). Ils sont désignés sous l'appellation « noms composés polylexicaux lâches »³². Dans les grandes lignes, on retrouve dans son étude les mêmes observations et analyses que chez Ch. Clairis et G. Babinotis (1996). Plus précisément, elle oppose le « nom composé polylexical lâche » au syntagme nominal : contrairement à ce dernier (exemple 1), la cohésion du « nom composé polylexical lâche » ne peut pas être interrompue par des éléments insérés dans la « portée » du figement, c'est-à-dire entre le premier et le deuxième composant (exemple 2) :

(1) Syntagme nominal

Αυτή είναι η ζώνη, όπως φαίνεται, της ιδιαίτερης αντοχής

C'est la **zone**, semble-t-il, **de résistance particulière**

(2) Nom composé polylexical lâche

**Αυτή είναι η ζώνη, όπως φαίνεται, της ασφαλείας*

*C'est la **ceinture**, semble-t-il, **de sécurité**

Elle considère ces séquences comme des noms composés puisqu'elles rentrent dans la définition du nom composé qu'elle adopte – elles sont constituées d'au moins deux bases lexicales (mots racines) (A. Ralli 2004 : 202). Cependant, elle les appelle « noms composés polylexicaux lâches » puisqu'elles présentent des propriétés morphosyntaxiques qui ne sont

³² Nous traduisons le terme grec *χαλαρά πολυλεκτικά σύνθετα* par le terme français *noms composés polylexicaux lâches*.

pas observées dans les noms composés soudés. Ces propriétés comprennent la flexion interne (par exemple, flexion du premier composant) et les phénomènes d'accord morphologique (par exemple, en genre, cas et nombre entre l'adjectif et le nom). Enfin, elle note que, souvent, le sens de ces séquences n'est pas le produit du sens ordinaire de leurs éléments constitutifs (A. Ralli 2004 : 202). Par exemple, *ψυχρός πόλεμος*/guerre froide est une « sorte d'ambiance politique » et non pas « une guerre sous une température froide ».

Dans le cadre de la théorie du Lexique-Grammaire (M. Gross 1975, 1981), un certain nombre de travaux ont été consacrés à l'étude des noms composés du grec moderne. La recherche d'E. Sklavounou (1993) se situe dans le domaine spécialisé du tennisme et porte sur l'étude contrastive des noms composés de structures *Adjectif + Nom*, *Nom + Nom au génitif* et *Nom + Nom-relative* en grec, français et anglais. Dans cette perspective, elle a élaboré un dictionnaire électronique trilingue sous forme de graphes d'automates finis, utilisable à des fins de traduction automatisée.

Z. Gavrilidou (1997) fait une étude comparée des noms composés de structure *Nom + Nom* (e.g. *ταξίδι-αστραπή*/voyage-éclair) en grec et en français. Elle étudie en profondeur les notions de nom composé et de figement, et procède également à l'analyse syntaxique, sémantique et pragmatique des séquences *Nom + Nom*.

Enfin, T. Kyriacopoulou *et al.* (2002) présentent les premiers travaux sur le dictionnaire électronique des noms composés en grec moderne. Les sujets qui y sont abordés portent sur la flexion et les différentes variantes des noms composés ainsi que sur le programme de génération automatique des formes fléchies à partir des formes canoniques.

1.1.1 Classification des noms composés en grec moderne

En ce qui concerne le grec moderne et pour ce qui est des auteurs précités (cf. Ch. Clairis et G. Babiniotis 1996, A. Ralli 2004, T. Kyriacopoulou *et al.* 2002), la classification des noms composés est le plus souvent fondée sur leur structure lexicale interne. Elle concerne uniquement les noms composés binaires et elle ne contient pas les noms composés de toutes les structures morphosyntaxiques (sauf dans T. Kyriacopoulou *et al.* 2002). Par exemple, les noms composés de structure *Nom + Préposition + Nom*; tels que *άδεια άνευ αποδοχών*/congés non-payés, n'y figurent pas. Ci-dessous les classes que l'on y retrouve :

Structures morphosyntaxiques	Exemples
<i>Adjectif + Nom</i>	<i>ψυχρός πόλεμος/guerre froide</i> <i> τρίτος κόσμος/tiers monde</i>
<i>Nom + Nom au génitif</i>	<i>άρμα μάχης/char de combat</i> <i>ζώνη ασφαλείας/ceinture de sécurité</i>
<i>Nom + Nom</i> ³³	<i>νόμος πλαίσιο/loi-cadre</i> <i>λέξη κλειδί/mot-clé</i>

Pour ce qui est de notre recherche, la classification des noms composés que nous proposons est également fondée sur leur structure lexicale interne. Nous prenons en compte non seulement les noms composés binaires, mais aussi leurs expansions, c'est-à-dire les noms composés complexes (ou surcomposés) (e.g. *ΝΑ:Γ Ν:Γ=*: *φόρος προστιθέμενης αξίας/taxe sur la valeur ajoutée*). Nous reviendrons plus en détail sur la classification que nous adoptons dans II, 3.2.

Dans le cadre de cette étude, nous adoptons la définition formelle mise au point par M. Silberztein (1990), selon laquelle nous distinguons les mots simples, qui sont des séquences de lettres comprises entre deux séparateurs, et les mots composés, qui sont des séquences incluant au moins deux mots simples et donc au moins un séparateur. Les séparateurs en grec moderne sont le blanc, le trait d'union, la virgule de liaison et l'apostrophe. Nous considérons donc les noms composés soudés comme des mots simples, puisque leur traitement automatique est identique à celui des mots simples. Nous n'examinons ici que les noms composés non-soudés ou multi-mots.

Nous adoptons également la dichotomie terminologique proposée par Z. Gavriilidou (1997 : 22) : « les notions de *composition* (qui concerne la forme) et de *figement* (qui concerne la syntaxe interne et la sémantique du composé) sont bien distinctes en grec ». Nous tenons à préciser que dans le cadre de la présente étude, dont le but est de fournir des ressources

³³ A. Ralli (2004 : 199) n'inclut pas les noms composés *Nom + Nom* dans la typologie qu'elle propose car elle estime qu'il n'y a pas suffisamment d'indices pour qu'ils soient considérés des « noms composés polylexicaux lâches ».

lexicales qui seront utilisées à l'analyse automatique des textes, les définitions que nous utilisons sont motivées par le souci de l'exploitation informatique des données.

Nous adoptons aussi la position de G. Gross (1996a : 48), qui souligne : « dès lors qu'on envisage la composition nominale comme un phénomène qui affecte les groupes nominaux, une première constatation s'impose : si l'on veut décrire ce phénomène de façon intégrée et reproductible, il est impossible de se contenter d'observations générales sur le figement, indépendamment de la nature de la structure interne des noms composés. En effet, pour reconnaître automatiquement les noms composés, il faut être en mesure d'en montrer les limites et prédire leur morphologie (pour le français : formation du pluriel et éventuellement du féminin), et établir au préalable une typologie de la composition, afin de percevoir les problèmes spécifiques à chaque type. En fonction de la structure interne des noms composés, les éléments de la description varient. Selon que la structure interne du mot composé correspond ou non à la structure canonique des groupes nominaux, les indications requises ne sont pas les mêmes. La reconnaissance automatique comme unité figée d'une structure hétérogène dans un texte est évidente, ce qui n'est pas le cas d'une structure canonique ».

Enfin, nous adoptons les analyses linguistiques de Ch. Clairis et G. Babinotis (1996), de A. Ralli (2004) et de A. Anastassiadis-Syméonidis (1986, 1996) et nous nous servons des critères qu'ils proposent sur la définition des noms composés. Toutefois, ces critères ne s'avèrent pas suffisants afin de définir et identifier les noms composés du grec moderne, au moins en ce qui concerne les $N(E+DET:G)N:G$. Ainsi, dans un but de raffinement, nous avons également eu recours aux études effectuées sur les noms composés du français et, notamment, aux travaux de G. Gross (1996a), A. Monceaux (1993) et A. Poncet-Montange (1991) (cf. I, 1.2).

1.2 Survol de la littérature sur les noms composés en français

Nous venons de présenter comment le phénomène de composition nominale est traité dans la littérature grammaticale grecque. Notre choix d'effectuer un survol de la littérature sur les noms composés du français aussi se justifie par le fait qu'on y retrouve des problèmes similaires (ce qui n'est pas le cas, par exemple, avec les noms composés de l'allemand qui présentent des particularités complètement différentes). Pour ce qui est du français, il existe également un nombre important de travaux portant sur la composition lexicale (ou morphologique) en général et, en particulier, sur les noms composés. Ces études sont issues soit de l'approche linguistique, soit de l'approche statistique, soit d'une approche hybride

combinant des éléments des deux premières approches. Dans ce sous-chapitre, nous citons certains travaux qui ont été effectués dans le cadre de l'approche linguistique et qui servent aujourd'hui de référence pour la plupart des études concernant la composition nominale en français.

Plusieurs auteurs ont déjà proposé des survols des travaux linguistiques sur les noms composés en français, parmi lesquels nous citons : A. Poncet-Montange 1991, A. Monceaux 1993, G. Gross 1996a et Z. Gavriilidou 1997³⁴. Dans ce sous-chapitre, nous ne reprenons que les grandes lignes de la réflexion sur la composition lexicale (ou morphologique) en français et, plus particulièrement, sur la composition nominale.

La première étude qui traite les mots composés du français de manière approfondie est le *Traité de la formation des noms composés* (A. Darmesteter 1874). L'auteur se propose d'étudier « les lois de la composition française », c'est-à-dire « les règles auxquelles elle est soumise et les procédés dont elle use lorsque, empruntant à la langue des mots déjà existants, elle les combine dans des créations nouvelles ». L'ouvrage propose, en réalité, un traitement sémantique pour les noms composés. Selon A. Darmesteter (1874), « le nom composé évoque dans l'esprit non les images distinctes répondant à chacun des mots composants mais une image unique ». L'auteur fait la distinction entre les « composés véritables » (e.g. *timbre-poste*, *rouge-gorge*), qui sont asyntaxiques car elliptiques, et les « composés apparents » (e.g. *pomme de terre*, *boîte à lettres*), qui sont conformes aux lois de la syntaxe et qui doivent leur existence aux hasards de l'usage.

Le critère purement sémantique de l'« unité d'image » a été repris par de nombreux linguistes dont Ch. Bally (1965), F. Brunot et Ch. Bruneau (1949) et M. Grevisse (1986). Toutefois, comme le note G. Gross (1986b : 54-55), « la notion d'*idée unique* qui correspond à celle de signifié unique ne concerne que les mots composés "figés", ceux dont aucun élément ne peut faire l'objet d'un choix : *un blanc-bec*, *une pièce rapportée*, *un chat-huant*, *un cordon bleu*, etc. Mais, il en est des noms composés comme des phrases figées : le figement peut affecter l'ensemble ou un seul des éléments constitutifs, avec toutes les combinaisons intermédiaires possibles. Or, les composés totalement figés constituent une minorité. Près de quatre composés sur cinq comprennent une position susceptible de permettre un choix ».

³⁴ Z. Gavriilidou (1997) a effectué une étude comparée en français et en grec.

La définition fonctionnaliste des mots composés est motivée par leur comportement syntaxique. A. Martinet (1967) propose le terme de « synthème » pour désigner « les unités linguistiques dont le comportement syntaxique est strictement identique à celui des monèmes avec lesquels elles commutent, mais qui peuvent être conçues comme formées d'éléments sémantiquement identifiables ». Comme le note A. Poncet-Montange (1991 : 7), « la véritable innovation de Martinet consiste à traiter le problème de la définition des mots composés en termes formels et à l'aide de critères objectifs ». Du point de vue syntaxique, les « synthèmes » se comportent exactement comme des mots simples (monèmes), avec lesquels ils sont commutables :

Luc dort sur (la chaise longue + le canapé)

Les éléments constitutants perdent leur autonomie syntaxique et ne peuvent pas être déterminés individuellement. Toute détermination porte obligatoirement sur l'ensemble du « synthème » :

* *Luc dort sur la chaise (plus + infiniment) longue*

* *Luc dort sur la chaise confortable longue*

Mais : *Luc dort sur une chaise longue confortable*

A. Martinet (1967) reconnaît qu'il existe des degrés de figement sémantique, mais il souligne aussi qu'« il est pratiquement impossible de tracer une limite précise entre ce qui est sémantiquement analysable et ce qui ne l'est pas ». La définition syntaxique est sans doute plus rigoureuse, mais elle s'avère elle-aussi insuffisante pour l'étude de l'ensemble des mots composés.

L'idée d'A. Darmesteter (1874) qu'un nom composé est « une phrase en réduction » a été reprise par des auteurs d'inspiration générativiste, comme par exemple L. Guilbert 1971. Selon l'hypothèse transformationniste (L. Guilbert 1971), les mots dérivés et les mots composés sont formés à partir de phrases de base sous-jacentes. Par exemple, les noms composés de structure *Nom + Adjectif* seraient issus d'une série de transformations portant sur une phrase de base en *être*. La relativation, ensuite l'effacement du pronom relatif et du verbe *être* en constituent les principales étapes :

N₀ être Adj = : Cette chaise est longue

[que être]=: cette chaise qui est longue

[Déf=UN]=: une chaise qui est longue

[que être z.]=: une chaise longue³⁵

G. Gross (1996a) présente une étude globale sur les noms composés et sur d'autres locutions telles que les déterminants composés, les verbes composés et les locutions verbales, les locutions adjectivales, les locutions adverbiales et les locutions prépositives et conjonctives. Il applique une série de critères afin de définir le phénomène linguistique de figement. Les propriétés qui caractérisent le figement comprennent, notamment, la polylexicalité, l'opacité sémantique, le blocage des propriétés transformationnelles, la non-actualisation des éléments constituants, le blocage des paradigmes synonymiques et la non-insertion³⁶. Pour ce qui est des noms composés, il en étudie les propriétés morphosyntaxiques et sémantiques spécifiques, et en propose une typologie de classification tout en focalisant sur l'analyse des noms composés de structures *Nom + Adjectif* et *Nom de Nom*.

Dans le cadre de la théorie du Lexique-Grammaire, de nombreuses études ont été effectuées sur les noms composés. A titre indicatif, citons les travaux de A. Monceaux (1993) sur les noms composés de structure *Nom + Adjectif*, de A. Poncet-Montagne (1991) sur les noms composés de structures *Nom à Nom* et *Nom à Verbe* ainsi que les travaux de A. Firenze et B. Pelletier (1989), G. Gross (1991), G. Gross et A. Dugas (1991) et de A. Belanger (1997) sur les noms composés de structure *Nom de Nom*.

Enfin, dans une perspective plus applicative visant à l'annotation des noms composés dans des corpus de textes écrits, E. Laporte ; T. Nakamura et S. Voyatzi (2008)³⁷ proposent une classification morphosyntaxique d'un ensemble de 125 791 noms composés (lemmes) du français. De nombreux problèmes y sont discutés concernant, notamment, la délimitation lexicale (ou morphologique) et syntaxique des noms composés, le traitement des noms composés prédicatifs, la représentation formalisée des formes coordonnées dues à la factorisation des parties communes ainsi que le problème de limites entre noms composés et entités nommées.

³⁵ Pour une présentation plus détaillée de l'hypothèse transformationniste de formation des noms composés, voir aussi : A. Poncet-Montagne 1991 et Z. Gavriilidou 1997.

³⁶ Pour une présentation plus détaillée de ces propriétés, cf. aussi I, 2.2.1-2.2.11.

³⁷ Désormais E. Laporte *et al.* (2008).

Chapitre 2. La notion de figement

2.1 Introduction

Dans la littérature linguistique, les constructions figées ou idiomatiques ont préoccupé plusieurs auteurs (pour le français : L. Danlos 1981 ; 1988, E. Laporte 1988, G. Gross 1996a, M. Gross 1982 ; 1985 ; 1988b ; 1988c ; 1996, J. Klein *et al.* 2003, J. Klein et B. Lamiroy 2005, B. Lamiroy 2003 ; 2006, M. Silberztein 1990 ; pour le grec : A. Fotopoulou 1990 ; 1993a, Z. Gavriilidou 1997, A. Moustaki 1995, O. Tsaknaki 2005, S. Voyatzi 2006).

La distinction entre mots composés et séquences libres (ou compositionnelles) est fort dépendante de la définition de la notion de figement (A. Anastassiadis-Syméonidis 1986). Notons, tout d'abord, que les noms composés ont la même distribution syntaxique que les noms simples. « Ils relèvent donc d'une analyse à deux niveaux : du point de vue externe, ils ont dans la phrase une fonction de groupe nominal et leur structure interne n'est pas pertinente. En revanche, si on les analyse au regard de leur constitution interne, on observe que ce sont des suites qui n'ont pas la liberté de fonctionnement des groupes nominaux ordinaires, ce qui revient à dire, en particulier pour le traitement informatique, qu'ils comportent des séparateurs (blancs, apostrophes ou traits d'union) qu'on ne doit pas interpréter comme des articulations. Les noms composés ont donc cette particularité qu'ils allient l'unité à la pluralité » (G. Gross 1996a : 27).

Dans ce chapitre, nous étudions la notion générale de figement tout en mettant l'accent sur les aspects particuliers que celle-ci peut revêtir dans les noms composés $N(E + DET:G)N:G$. Nous donnons également un récapitulatif des propriétés générales du figement qui nous serviront de critères de délimitation des noms composés $N(E + DET:G)N:G$.

Deux types de critères motivent la délimitation des noms composés $N(E + DET:G)N:G$. D'une part, le figement est d'abord et traditionnellement une notion intuitive que l'on résume habituellement par la formule : « le sens d'un mot composé ne se déduit pas du sens de ses composants. La recherche de critères de définition plus objectifs que l'intuition sémantique a conduit à rendre compte de phénomènes linguistiques de nature très diverse mais qui ne sont pas indépendants les uns des autres, décrits par le terme de figement » (G. Gross 1996a : 9). En d'autres termes, il s'agit d'associer l'intuition sémantique à certaines restrictions

observables dans le comportement syntaxique des composants. Cette recherche concerne les mots composés « syntaxiques », c'est-à-dire les mots composés qui ne se distinguent pas formellement de syntagmes libres ordinaires tels que, ici en l'occurrence, les noms composés de structure $N(E + DET:G) N:G$. Comme le signale G. Gross (1996a : 7), « on est en présence de deux paramètres différents : d'une part le fonctionnement syntaxique interne d'une suite donnée, qui peut être libre ou faire l'objet de différents niveaux de restrictions, et, d'autre part, la signification de la suite qui est ou non le produit de celle des éléments constitutifs ». Nous nous rendons donc compte que le figement sémantique et le figement syntaxique sont les deux aspects d'un même phénomène.

G. Gross (1990 ; 1996a) propose une série de critères syntaxiques et sémantiques qui permettent de distinguer les formes figées des formes libres. Rappelons ici le principe général de cette distinction : « Une séquence de mots simples est figée (ou composée) si l'une au moins de ses propriétés syntaxiques, distributionnelles ou sémantiques ne peut pas être déduite des propriétés de ses constituants ». Dans le cadre de cette étude, nous adoptons les critères proposés par G. Gross (1996a : 9-22), tout en essayant de les adapter aux $N(E + DET:G) N:G$ du grec moderne (cf. I, 2.2.1-2.2.11).

2.2 Les critères généraux et les propriétés générales du phénomène de figement

Les critères généraux (ou propriétés générales) qui mettent en évidence le phénomène de figement s'appliquent au niveau lexical, syntaxique et sémantique. De manière générale, nos critères de figement se réfèrent à des tests bien précis, auxquels nous avons soumis les séquences de structure $N(E + DET:G) N:G$. En revanche, les propriétés de figement ne se réfèrent à aucun test, mais résultent d'observations qui nous aident souvent à mesurer le degré de figement. Nous présentons ci-dessous les critères que nous avons appliqués aux séquences de structure $N(E + DET:G) N:G$ du grec moderne et qui prennent appui sur ceux proposés par G. Gross (1996a : 9-22). Les propriétés de figement sont présentées à la suite des critères. Notons que les critères sont exposés tout au long de ce chapitre selon l'ordre de leur application, à savoir :

- la polylexicalité (ou la combinatoire lexicale),

- la non-compositionnalité du sens ou les contraintes sémantiques³⁸,
- le blocage des propriétés transformationnelles (ou les contraintes syntaxiques),
- la non-actualisation des éléments constitutifs,
- la non-prédication des noms composés,
- le blocage des paradigmes synonymiques (ou les contraintes lexicales),
- la non-insertion,
- la coordination,
- le degré de figement.

Les deux propriétés que nous avons étudiées et qui sont étroitement liées aux critères de figement sont les suivantes :

- le défigement,
- l'étymologie.

2.2.1 La polylexicalité

« La première condition nécessaire pour qu'on puisse parler de figement est que l'on soit en présence d'une séquence de plusieurs mots et que ces mots aient, par ailleurs, une existence autonome. Cela exclut les suites formées à l'aide d'un affixe (*i.e.* préfixe, suffixe), qui relèvent de ce qu'on appelle la dérivation » (G. Gross 1996a : 9). Comme nous l'avons signalé dans I, 1.1.1, du point de vue du TAL, la séparation entre mots simples et mots composés (ou « multi-mots ») est purement graphique : un mot composé (ou « multi-mot ») est une suite constituée d'au moins deux mots simples et un séparateur (M. Silberztein 1990). Nous admettons comme séparateurs des noms composés du grec moderne les séparateurs suivants : le blanc, le trait d'union et l'apostrophe³⁹ :

ζώνη ασφαλείας

³⁸ Gross (1996 : 22) propose d'étudier l'opacité sémantique en tant que propriété générale du figement.

³⁹ Notons que, selon les règles typographiques du grec moderne, l'apostrophe est toujours suivie d'un blanc. Cependant, dans la pratique, il arrive très souvent que l'apostrophe ne soit pas suivie de blanc dans les textes écrits. Pour ce qui est de cette étude, nous tenons compte des deux écritures.

(ceinture de sécurité)

δακτυλικό αποτύπωμα

(empreinte digitale)

νόμος-πλαίσιο

(loi cadre)

εισιτήριο μετ' επιστροφής

(billet aller retour)

2.2.2 La non-compositionnalité du sens (ou les contraintes sémantiques)

La notion de compositionnalité repose traditionnellement sur le fait que « le sens d'une séquence donnée est le produit du sens de ses composants. Ainsi le sens d'une phrase est facteur de celui de son prédicat et de celui de ses arguments » (G. Gross 1996a : 10). Pour illustrer la notion de compositionnalité, prenons l'exemple de la phrase suivante :

Η Μαρία διαβάζει ένα βιβλίο

La-*Nfs* Maria-*Nfs* lit un-*Gns* livre-*Gns*

(Marie lit un livre)

Le sens de cette phrase est compositionnel, car il se déduit de celui de son prédicat et de celui de ses arguments. Cependant, dans les langues il existe également un nombre important de phrases qui ne sont pas compositionnelles. Ces phrases ne peuvent pas être interprétées littéralement, car les mots qui les composent n'y conservent pas leur sens habituel. Voici un exemple d'une phrase non-compositionnelle :

Μου ανεβαίνει το αίμα στο κεφάλι

Me-*Gs* monte le-*Nns* sang-*Nns* à la-*Ans* tête-*Ans*

(La moutarde me monte au nez)

« Ce que nous venons de dire de la phrase s'applique également aux unités de niveau inférieur. Un groupe nominal s'interprète en fonction du sens ordinaire de ses éléments constitutifs : substantif-tête et modifieur, que ce modifieur soit une relative, un adjectif ou un complément de nom » (G. Gross 1996a : 11). Ceci est donc valable pour les séquences de

structure $N (E + DET:G) N:G$ que nous étudions dans le cadre de ce travail. Ainsi, dans l'exemple :

<i>το βιβλίο της Μαρίας</i>	(=: <i>το βιβλίο που ανήκει στη Μαρία</i>)
(le livre de Marie)	(=: le livre qui appartient à Marie)

le sens du groupe nominal est compositionnel, c'est-à-dire qu'il se déduit du sens ordinaire de ses éléments constitutifs. En revanche, dans l'exemple suivant :

<i>το κουτί της Πανδώρας</i>	(≠: <i>το κουτί που ανήκει στην Πανδώρα</i>)
(la boîte de Pandore)	(≠: la boîte qui appartient à Pandore)

le sens du groupe nominal n'est pas compositionnel, car on ne parle pas d'« une boîte qui appartient à Pandore », mais d'« une situation qui constitue la source de nombreux malheurs ». A l'instar des phrases qui donnent lieu à une double lecture (cf. G. Gross 1996a : 11), il existe également des groupes nominaux qui donnent lieu à deux lectures, une compositionnelle et une non-compositionnelle. Par exemple le groupe nominal *πρίγκιπας του παραμυθιού*/prince de la fable peut être interprété de manière compositionnelle et, dans ce cas, désigner le protagoniste d'une histoire imaginaire. Le même groupe nominal peut donner lieu à une lecture non-compositionnelle :

πρίγκιπας του παραμυθιού
(prince charmant)

Dans le deuxième cas, il ne se réfère pas au protagoniste d'une histoire imaginaire, mais il désigne l'homme idéal. Dans ce cas, nous sommes en présence d'une suite sémantiquement figée et contrainte lexicalement⁴⁰. Notons aussi que, selon E. Laporte (1988 : 121), « même lorsqu'on peut employer une locution dans le sens littéral, on a tendance à éviter de le faire, car l'interprétation idiomatique est préférée à l'interprétation littérale ».

Les exemples que nous avons examinés jusqu'à présent représentent les deux extrêmes du phénomène de figement. Il s'agit soit de séquences compositionnelles (ou libres) soit de séquences qui mettent en jeu un figement qu'on pourrait appeler « total », puisque aucun des éléments de la séquence ne permet de choix ni au niveau lexical ni au niveau syntaxique. Cependant, comme nous le montrerons par la suite (cf. I, 2.2.9), la non-compositionnalité du

⁴⁰ Sur les contraintes lexicales cf. I, 2.2.6.

sens est un phénomène scalaire : le degré de figement est un paramètre qui calcule les contraintes de nature lexicale, syntaxique et sémantique, affectant la relation entre les éléments constitutifs d'une suite donnée.

2.2.3 Le blocage des propriétés transformationnelles

Les constructions libres ont des propriétés transformationnelles qui dépendent de leur organisation interne. En ce qui concerne les verbes ou, plus précisément, la relation verbe-complément(s) (V-W), nous observons des transformations régulières telles que la passivation, la pronominalisation, la relativation, le détachement etc. Par exemple, pour la phrase :

Η Μαρία έγραψε ένα ποίημα

(Marie a écrit un poème)

les transformations suivantes sont possibles :

- passivation :

Αυτό το ποίημα γράφτηκε από τη Μαρία

(Ce poème a été écrit par Marie)

- pronominalisation :

Η Μαρία το έγραψε

(Marie l'a écrit)

- relativation :

Το ποίημα που έγραψε η Μαρία [είναι συγκινητικό]

(Le poème que Marie a écrit [est émouvant])

- détachement :

Αυτό το ποίημα, το έγραψε η Μαρία

(Ce poème, Marie l'a écrit)

« Toutes ces modifications ne s'appliquent pas de façon systématique à l'ensemble des relations verbe-compléments. Sur les quelques dizaines de transformations habituelles, on peut observer l'absence de telle ou telle, dont il n'est pas toujours facile de percevoir la

cause » (G. Gross 1996a : 12). Le figement ne s'appliquera donc qu'aux structures qui se caractérisent par le blocage de propriétés transformationnelles, autrement dit par des contraintes syntaxiques.

La remarque que nous venons de faire au sujet des suites verbales s'applique également aux suites $N(E + DET:G)N:G$ que nous étudions ici. Diverses transformations, telles que la nominalisation (voir ci-dessous exemples i, et iv), l'adjonction de modifieurs (exemples ii et v) et la prédicativité (exemples iii et vi), qui sont acceptables dans le cas des syntagmes nominaux libres, sont interdites pour les syntagmes nominaux figées (ou noms composés). En guise d'exemple, citons la séquence nominale libre ci-dessous :

υπολογιστής αξιοπιστίας
(ordinateur de fiabilité)

- (i) *η αξιοπιστία αυτού του υπολογιστή*
(la fiabilité de cet ordinateur)
- (ii) *ένας υπολογιστής μεγάλης αξιοπιστίας*
(un ordinateur de grande fiabilité)
- (iii) *Αυτός ο υπολογιστής είναι αξιόπιστος*
(Cet ordinateur est fiable)

En revanche, ces transformations ne sont pas autorisées dans le cas d'une séquence syntaxiquement figée :

πνεύμονας πρασίνου
(poumon vert)

- (iv) **το πράσινο αυτού του πνεύμονα*
(*le vert de ce poumon)
- (v) **ένας πνεύμονας μεγάλου πρασίνου*
(*un poumon de grand vert)
- (vi) **Αυτός ο πνεύμονας είναι πράσινος*
(*Ce poumon est vert)

2.2.4 La non-actualisation des éléments constitutifs

Une condition supplémentaire qui nous permettra de parler de figement ou de degrés de figement est la non-actualisation des éléments constitutifs d'une séquence. A la différence des groupes nominaux libres, dont chaque élément lexical peut recevoir une actualisation (détermination ou modification) autonome, les noms composés reçoivent une actualisation globale, qui porte sur l'ensemble du groupe nominal et non sur l'un ou l'autre des composants (cf. G. Gross 1996a : 32). On opposera donc le groupe nominal libre *το βιβλίο του φίλου μου*/le livre de mon ami, dont les éléments constitutifs peuvent être actualisés indépendamment (exemples i, ii, iii, iv et v), au nom composé *μερίδα του λέοντος*/part du lion, où l'actualisation ne s'applique ni au nom tête (exemple x) ni au deuxième composant nominal (exemples vi, vii et ix). Dans le premier cas, les deux noms peuvent être actualisés :

το βιβλίο του φίλου μου

le-*Nns* livre-*Nns* le-*Gms* ami-*Gms* moi-*Gms*

(le livre de mon ami)

- (i) *το βιβλίο αυτού του φίλου μου* (DET=Dém)
le-*Nns* livre-*Nns* ce-*Gms* le-*Gms* ami-*Gms* moi-*Gms*
(le livre de cet ami à moi)
- (ii) *το βιβλίο ενός φίλου μου* (DET=UN)
le-*Nns* livre-*Nns* un-*Gms* ami-*Gms* moi-*Gms*
(un livre d'un ami à moi)
- (iii) *Αυτό το βιβλίο είναι του φίλου μου* (prédicativité)
Ce-*Nns* le-*Nns* livre-*Nns* est le-*Gms* ami-*Gms* moi-*Gms*
(Ce livre est de mon ami)
- (iv) *το βιβλίο του καλύτερου φίλου μου* (Modif N:G)
le-*Nns* livre-*Nns* le-*Gms* meilleur-*Gms* ami-*Gms* moi-*Gms*
(le livre de mon meilleur ami)
- (v) *το υπέροχο βιβλίο του φίλου μου* (Modif N)
le-*Nns* formidable-*Nns* livre-*Nns* le-*Gms* ami-*Gms* moi-*Gms*
(le livre formidable de mon ami)

En revanche, dans l'exemple :

μερίδα του λέοντος

part le-*Gms* lion-*Gms*

(part du lion)

aucun des éléments constitutifs ne peut recevoir d'actualisation propre :

(vi) **μερίδα αυτού του λέοντος* (DET=Dém)

*part ce-*Gms* le-*Gms* lion-*Gms*

(*part de ce lion)

(vii) **μερίδα ενός λέοντος* (DET=UN)

*part un-*Gms* lion-*Gms*

(*part d'un lion)

De plus, il ne peut pas y avoir de relation prédicative entre *μερίδα*/partage et *λέων*/lion :

(viii) **Αυτή η μερίδα είναι του λέοντος* (prédicativité)

*Cette-*Nfs* part-*Nfs* est le-*Gms* lion-*Gms*

(*Cette part est du lion)

Les éléments constitutifs ne peuvent pas recevoir de modifieurs non plus. Par exemple :

(ix) **μερίδα του άγριου λέοντος* (Modif *N:G*)

*part le-*Gms* lion-*Gms* sauvage-*Gms*

(*part du lion sauvage)

(x) **μεγάλη μερίδα του λέοντος* (Modif *N*)

*grande part le-*Gms* lion-*Gms*

(*grande part du lion)

G. Gross (1996a : 33) note aussi que le déterminant du deuxième composant peut constituer un indice de figement : dans les noms composés mettant en jeu deux substantifs, le second substantif a l'article zéro. Par exemple : *NN:G*=: *ζώνη ασφαλείας*/ceinture de sécurité.

Selon G. Gross (1996a : 33) et A. Anastasiadis-Syméonidis (1986), le seul autre déterminant qui peut apparaître devant le deuxième composant nominal est le générique *le*⁴¹ :

⁴¹ A noter que les entités nommées se différencient sur ce point des noms composés communs.

N DET:G N:G
μαλλί της γριάς
 cheveu la-*Gfs* vieille-*Gfs*
 (barbe à papa)

Dans ce cas, « il y a bien ici présence d'un article, mais on observe qu'il ne réfère à aucun objet ou être défini, qu'il est totalement contraint et ne peut faire l'objet d'aucune substitution » :

**μαλλί γριάς* (DET=E)
 cheveu vieille-*Gfs*
 (barbe à papa)

**μαλλί μιας γριάς*⁴² (DET=UN)
 cheveu une-*Gfs* vieille-*Gfs*
 (*barbe à un papa)

**μαλλί αυτής της γριάς* (DET=Ddém)
 cheveu cette-*Gfs* vieille-*Gfs*
 (*barbe à ce papa)

**μαλλί της γριάς μας* (DET=Poss)
 cheveu la-*Gfs* vieille-*Gfs* notre-*Gfs*
 (*barbe à notre papa)

2.2.5 La non-prédication des noms composés

« Le fait que dans le nom composé aucun élément ne puisse être actualisé montre qu'il n'est pas le siège d'une prédication. Un *fait évident* est une prédication, un *fait divers* ne l'est pas : *Nous avons constaté un fait qui est évident*, **Nous avons constaté un fait qui est divers*. Cela veut dire que le groupe nominal ordinaire est une assertion que le locuteur est en mesure de faire suivant les règles de la grammaire, tandis que le nom composé est une séquence qui réfère à un objet ou une idée abstraite que le locuteur ne crée pas, mais qui est préconstruite et

⁴² Notons que la lecture compositionnelle est également possible pour les exemples *το μαλλί μιας γριάς*/les cheveux d'une vieille dame, *το μαλλί αυτής της γριάς*/les cheveux de cette vieille dame, *το μαλλί της γριάς μας*/les cheveux de notre vieille dame. Sur ce point, voir aussi I, 2.2.2.

qui fait partie de son stock lexical au même titre que les noms simples » (G. Gross 1996a : 33).

Le phénomène de la non-prédication s'observe aussi aux noms composés de structure $N(E + DET:G) N:G$. Ainsi dans le cadre d'un groupe nominal ordinaire de structure $N(E + DET:G) N:G$:

η ζώνη της αδελφής μου

(la ceinture de ma sœur)

on peut avoir la prédication :

Αυτή η ζώνη είναι της αδελφής μου

(Cette ceinture est de ma sœur)

En revanche, dans le cas d'un nom composé de la même structure :

ζώνη ασφαλείας

(ceinture de sécurité)

la prédication n'est pas possible :

**Αυτή η ζώνη είναι ασφαλείας*

(*Cette ceinture est de sécurité)

2.2.6 Le blocage des paradigmes synonymiques (ou les contraintes lexicales)

Depuis la linguistique saussurienne, on oppose deux axes dans le système langagier : l'axe syntagmatique et l'axe paradigmatique. L'axe paradigmatique veut qu'en position d'arguments on ait affaire non à des unités mais à des classes de mots. Ainsi, dans les structures libres, il est possible de remplacer un mot soit par un autre mot de la même classe sémantique soit par un synonyme. Ces possibilités de substitution synonymique dépendent de la nature des prédicats.

Nous observons que dans les suites figées de structure $N(E + DET:G) N:G$ cette possibilité de substitution est interdite ou, au moins, assez limitée (cf. G. Gross 1996a : 18). Par exemple, on peut avoir :

προσομοιωτής πτήσης

simulateur vol-*Gfs*
(simulateur de vol)

αδελφή του ελέους
sœur le-*Gms* miséricorde-*Gms*
(sœur de la miséricorde)

απεργία ζήλου
grève zèle-*Gms*
(grève du zèle)

mais on ne peut pas avoir :

*μιμητής πτήσης (N synonyme)
*imitateur vol-*Gfs*
(*imitateur de vol)

*αδελφή του οίκτου (N:G synonyme)
*sœur la-*Gms* pitié-*Gms*
(*sœur de la pitié)

*αδελφή της ευσπλαχνίας (N:G parasynonyme)
*sœur la-*Gms* compassion-*Gms*
(*sœur de la compassion)

*απεργία προθυμίας (N:G parasynonyme)
*grève ardeur-*Gfs*
(*grève d'ardeur)

Il existe aussi des cas où les composants lexicaux des séquences *N (E + DET:G) N:G* n'ont pas d'existence autonome en dehors de la séquence et, par conséquent, toute sorte de variation lexicale y est interdite. Nous parlons, à leur propos, des « hapax lexicaux » (M. Gross 1990 : 182). A titre indicatif, citons les exemples suivants :

παρανάλωμα ((E + του) πυρός + (E + της) φωτιάς)⁴³

⁴³ Cette séquence fait partie de la phrase simple figée *ΝΟ γίνομαι παρανάλωμα (του πυρός + της φωτιάς)*/être dévorée par les flammes, mais nous la prenons quand même en compte ici, car elle peut avoir un emploi

(tout a été mis dans le feu)

γέεννα του πυρός

(géhénne)

Les noms *παρανάλωμα*/feu et *γέεννα*/géhénne n'apparaissent que dans les séquences ci-dessus et, par conséquent, ne peuvent pas être substitués par d'autres noms.

Comme le note S. Voyatzi (2006 : 72) pour les adverbess figés, un autre phénomène propre à la langue grecque est à remarquer. Il s'agit de la possibilité de substitution des séquences, dont les composants lexicaux appartiennent à des systèmes de langues marginaux, à savoir le grec savant (ou « katharevoussa ») et le grec ancien, par leurs équivalentes sémantiques du grec populaire (ou « démotique »). Le même phénomène s'observe dans des séquences *N (E + DET:G) N:G*. Notons que, dans le cadre de la présente étude, nous appelons les structures en grec savant ou ancien en général *variantes savantes*.

Cette possibilité de substitution sera étudiée en détail dans II, 5.2.1.1 et 5.3.1.1, il n'est donc pas nécessaire d'y insister plus ici. Nous nous contenterons simplement d'en citer quelques exemples à titre indicatif :

- les *variantes savantes* peuvent être substituées par les *variantes démotiques* indépendamment du prédicat de la phrase où ils apparaissent :

παρανάλωμα του πυρός (variante savante)

παρανάλωμα της φωτιάς (variante démotique)

(feu)

αγιασμός των υδάτων (variante savante)

αγιασμός των νερών (variante démotique)

(bénédictiion des eaux)

μάρτυς κατηγορίας (variante savante)

μάρτυρας κατηγορίας (variante démotique)

(témoin à charge)

- les *variantes savantes* ne peuvent pas être substituées par les *variantes démotiques* :

<i>κέρας της Αμαλθείας</i>	(variante savante)
<i>*κέρατο της Αμάλθειας</i>	(variante démotique)
(corne d'Amalthée)	
<i>βάπτισμα του πυρός</i>	(variante savante)
<i>*βάπτισμα της φωτιάς</i>	(variante démotique)
(baptême du feu)	
<i>λάμπα θυέλλης</i>	(variante savante)
<i>*λάμπα θύελλας</i>	(variante démotique)
(lampe-tempête)	

Nous pourrions donc conclure qu'une distinction entre séquences figées $N (E + DET:G) N:G$ et séquences libres de la même structure, qui serait essentiellement fondée sur la présence ou l'absence de variantes lexicales, ne s'avérerait pas vraiment opératoire pour le grec moderne. Toutefois, l'existence de variantes lexicales nous servira de critère pour mesurer le degré de figement des structures $N (E + DET:G) N:G$ ici étudiées (cf. I, 2.2.9).

2.2.7 La non-insertion

Dans les suites libres, il est possible d'insérer des modifieurs, comme par exemple un adjectif (exemple i) ou une relative (exemple ii) dans le groupe nominal, un adverbe d'intensité avant l'adjectif (exemple iii) etc. (cf. G. Gross 1996a : 18). A titre indicatif, citons les exemples suivants :

- Η Μαρία διαβάζει ένα βιβλίο*
(Maria lit un livre)
- (i) *Η Μαρία διαβάζει ένα μεγάλο βιβλίο*
(Maria lit un grand livre)
- (ii) *Η Μαρία διαβάζει ένα βιβλίο που της αρέσει*
(Maria lit un livre qu'elle aime)
- (iii) *Η Μαρία διαβάζει ένα πολύ μεγάλο βιβλίο*
(Maria lit un très grand livre)

στρέμματα/Le feu a détruit 1000 hectares. Sur ce point, cf. aussi III, 6.8.

Dans les séquences figées, l'insertion des modifieurs est interdite ou réduite. En particulier, en ce qui concerne les séquences figées $N (E + DET:G) N:G$, nous avons observé qu'on ne peut pas insérer des modifieurs adjectivaux dans la « portée » du figement. Il est possible de qualifier une telle séquence au moyen d'un adjectif, mais cet adjectif sert à qualifier l'ensemble de la séquence figée et non seulement le composant nominal devant lequel l'adjectif est inséré. Notons que l'insertion d'un adjectif n'est pas autorisée pour le deuxième composant nominal non plus. A titre indicatif citons les exemples suivants :

βάση δεδομένων

base données-*Gnp*

(base de données)

μεγάλη βάση δεδομένων

grande base données-*Gnp*

(grande base de données)

**βάση μεγάλων δεδομένων*

*base grandes-*Gnp* données-*Gnp*

(*base de grandes données)⁴⁴

πύργος ελέγχου

tour contrôle-*Gms*

(tour de contrôle)

καινούριος πύργος ελέγχου

nouvelle tour contrôle-*Gms*

(nouvelle tour de contrôle)

**πύργος απόλυτου ελέγχου*

*tour absolu-*Gms* contrôle-*Gms*

(*tour de contrôle absolu)

⁴⁴ Notons qu'une suite telle que *βάση ορολογικών δεδομένων*/base terminologiques-*Gnp* données-*Gnp*/(base de données terminologiques) est acceptable. Dans ce cas, nous n'avons pas affaire à l'insertion d'un modifieur adjectival, mais à un nom composé complexe (ou surcomposé) de structure $N A:G N:G$. Les noms composés complexes (ou surcomposés) sont étudiés dans III, 6.3.

2.2.8 La coordination

La coordination des constituants du groupe nominal est autorisée dans les séquences libres (G. Gross 1996a : 56). Par exemple, pour ce qui est des structures $N (E + DET:G) N:G$:

το σπίτι του Νίκου και της Μαρίας

(la maison de Nicolas et de Marie)

υιοθέτηση πολιτικής και στρατηγικής

(adoption de politique et de stratégie)

Il existe également des cas de séquences figées où la coordination est autorisée. Ceci est possible lorsque l'on a affaire à un nom composé comportant des modifieurs coordonnés. Dans ce cas, l'ensemble de la coordination fait partie de la séquence figée, ce qui donne lieu à des formes complexes figées. Par exemple :

αγαθά του Αβραάμ και του Ισαάκ

(biens d'Abraham et d'Isaac) (traduction littérale)

δρόμος της Αρετής και της Κακίας

(chemin de la vertu et de la méchanceté) (traduction littérale)

δικαίωμα του εκλέγειν και του εκλέγεσθαι

(droit d'élire et d'être élu)

μάχη θεών και γιγάντων

(combat de dieux et de géants)

Dans le cadre des $N (E + DET:G) N:G$, un autre phénomène de coordination est observé. Une séquence semi-figée peut être coordonnée avec une autre séquence semi-figée quand le nom-tête est le même dans les deux séquences. Il paraît que c'est une coordination des modifieurs en surface mais, en réalité, il s'agit d'entrées indépendantes. En fait, il s'agit du phénomène de factorisation, qui met en jeu des réductions des parties communes dans des formes coordonnées ou juxtaposées, problème ardu pour les analyseurs automatiques (cf. E. Laporte *et al.* 2008). Par exemple :

κρέμα ημέρας και κρέμα νυκτός=: *κρέμα ημέρας και νυκτός*⁴⁵

(crème de jour et crème de nuit=: crème de jour et de nuit)

γλυκά ταψιού και γλυκά κουταλιού=: *γλυκά ταψιού και κουταλιού*

(gâteaux de plateau et gâteaux de cuiller=: gâteaux de plateau et de cuillère)

ιστορία αγάπης και ιστορία μυστηρίου=: *ιστορία αγάπης και μυστηρίου*

(histoire d'amour et histoire de mystère=: histoire d'amour et de mystère)

Il s'agit d'un phénomène très courant en terminologie, difficile à résoudre dans le cadre du traitement automatique des textes.

Toutefois, cette caractéristique constitue un indice de figement. Nous observons que la coordination des modifieurs en surface n'est pas possible dans le cas de deux séquences entièrement figées ayant le même nom-tête :

**ζώνη ασφαλείας και αγνότητας*

(*ceinture de sécurité et de chasteté)

De plus, dans le cas où la coordination des constituants n'est pas autorisée, l'élément coordonné ne fait pas partie du nom composé :

**σιγή ιχθύος και πτηνού*

(*silence de mort et de décès)

**μερίδα του λέοντος και της τίγρης*

(*part du lion et du tigre)

Nous pourrions donc conclure que le critère de coordination ne s'avère pas vraiment opératoire pour le grec moderne. Toutefois, il nous sert de critère pour mesurer le degré de figement des séquences de structure $N (E + DET:G) N:G$ (cf. I, 2.2.9).

2.2.9 Le degré de figement

La non-compositionnalité du sens est un phénomène scalaire : elle peut être totale (e.g. *πέτρα του σκανδάλου*/pierre de scandale, *σανίδα σωτηρίας*/planche de salut), partielle (e.g. *σαλάμι*

⁴⁵ Cependant, la possibilité de l'existence d'un tel produit, et, par conséquent, d'un tel nom composé n'est pas exclue.

αέρος/saucisson de montagne, γλυκό του κουταλιού/gâteau à la cuillère) ou inexistante (e.g. κλαδί δέντρου/branche d'arbre) (cf. G. Gross 1996a : 11). Le figement peut donc être « complet » quand aucun des éléments composants de la séquence ne permettrait de choix et ne pourrait donc faire l'objet d'un paradigme. Dans le cas des exemples suivants, nous observons que les composants nominaux ne peuvent être substitués à aucun autre :

πέτρα του σκανδάλου

pierre le-G_{ns} scandale-G_{ns}

(pierre de scandale)

**λίθος του σκανδάλου*

(*N* savant)

**caillou le-G_{ns} scandale-G_{ns}*

(**caillou de scandale*)

**πέτρα του επεισοδίου*

(*N:G* parasyndrome)

**pierre le-G_{ns} incident-G_{ns}*

(**pierre de l'incident*)

μήλο της έριδος

pomme la-G_{fs} discorde-G_{fs}

(pomme de discorde)

**πορτοκάλι της έριδος*

(*N=: Nalim*)

**orange la-G_{fs} discorde-G_{fs}*

(**orange de discorde*)

**μήλο της διαφωνίας*

(*N:G* démotique)

**pomme le-G_{fs} conflit-G_{fs}*

(**pomme de conflit*)

σανίδα σωτηρίας

planche salut-G_{fs}

(planche de salut)

**σανίδα διάσωσης*

(*N* synonyme)

**planche sauvetage-G_{fs}*

(*planche de sauvetage)

*σανίδι σωτηρίας

(N neutre)

*planche salut-Gfs

(*barre de salut)

« Ces suites fonctionnent de façon compacte, en bloc, exactement comme les catégories simples (à des problèmes de graphie près) et c'est à tort ou pour des raisons de commodité rédactionnelle qu'on les fait figurer sous l'un ou l'autre des termes dans les dictionnaires. Elles devraient constituer des entrées indépendantes » (G. Gross 1996a : 16).

Cependant, les cas de figement total ne sont pas les plus fréquents. Dans la majorité des cas, nous avons affaire à des séquences qui autorisent une possibilité de paradigme. Le figement est donc associé à l'existence d'une restriction de sélection sémantique qui touche au moins à l'un des éléments constitutifs d'une séquence donnée. Examinons l'exemple suivant :

ειρωνεία της (τύχης + μοίρας)

ironie (de la chance + du destin)

Dans cette séquence, on peut remplacer *τύχη*/chance par *μοίρα*/destin. Il y a là une liberté lexicale, même si le sens reste non-compositionnel dans les deux cas.

Examinons maintenant les suites qui sont présentées dans les exemples suivants :

σαλάμι (αέρος + μπίρας)

(saucisson (de montagne + de bière))

γλυκό του (κουταλιού + ταψιού)

(gâteau (de la cuillère + du plateau))

Ces suites ne sont pas entièrement figées. Le deuxième composant nominal désigne respectivement différents types de charcuterie ou de gâteaux. D'ailleurs, le test de la règle d'identité met en évidence (ou non) l'emploi et le sens figés du nom tête de la suite :

Το σαλάμι αέρος είναι σαλάμι

(Le saucisson de montagne est un saucisson)

Το γλυκό του κουταλιού είναι γλυκό

Le gâteau à la cuillère est un gâteau

Mais :

**Η σανίδα σωτηρίας είναι σανίδα*

(*La planche de salut est une planche)

**Η πέτρα του σκανδάλου είναι πέτρα.*

(*La pierre de scandale est une pierre)

Le paradigme restreint sur l'un des deux éléments constitutifs peut être considéré comme un indice d'un degré non élevé de figement. Comme le souligne G. Gross (1996a : 17), « il existe donc des degrés de figement dans les langues, un continuum entre les séquences libres et celles qui sont entièrement contraintes » (cf. aussi M. Gross 1977, 1990). Ainsi, le degré de figement d'une structure sera décidé à partir des valeurs positives ou négatives d'un certain nombre de propriétés (ou critères) présentées précédemment.

2.2.10 Le défigement

Pour introduire la notion de « défigement », G. Gross (1996a : 19) confronte le comportement des constructions libres à celui des constructions figées : « les constructions libres sont caractérisées par l'existence de paradigmes permettant des substitutions définies par les contraintes d'arguments et par des modifications et des restructurations qui dépendent de la nature sémantique et syntaxique de la relation existant entre le prédicat et ses arguments. On peut ainsi calculer le nombre de variations potentielles pour une construction donnée. Toute transgression à ces possibilités est considérée comme une faute ».

En revanche, les constructions figées n'offrent pas cette possibilité. « Le figement peut être mis en évidence grâce à l'effet provoqué par le jeu du défigement, qui consiste à briser le carcan qui caractérise les séquences figées. Le défigement consiste à ouvrir des paradigmes là où, par définition, il n'y en a pas » (G. Gross 1996a : 20). C'est un phénomène que l'on observe, notamment, dans la presse, dans la publicité et dans les mots croisés dans le but d'attirer l'attention du lecteur ou du public. L'effet de surprise attendu met en évidence le phénomène du figement.

Le défigement peut porter sur des groupes nominaux et, ici en l'occurrence, sur les noms composés de structure $N(E + DET:G)N:G$. Citons ici quelques exemples de défigement, issus de la presse et du Web :

γιοφύρι της Άρτας

(pont d'Arta)

γιοφύρι (του περιφερειακού⁴⁶ + του Σταδίου Καραϊσκάκη⁴⁷)

(pont du périphérique + du Stade Karaïskaki)

ασκοί του Αιόλου

(outré d'Eole)

ασκοί (της Έλλης⁴⁸ + της κατασκοπίας⁴⁹ + της μισαλλοδοξίας⁵⁰)

(outré (d'Ellis + de l'espionnage + de l'intolérance))

2.2.11 Etymologie

A propos de l'origine du figement, G. Gross (1996a : 21) met au point la notion d'« étymologie » en tant que propriété générale du figement. Il remarque que « les constructions librement générées par la syntaxe et qui ne figurent, à plus forte raison, dans aucun dictionnaire » ne posent aucun problème au niveau de leur compréhension, même quand elles sont prononcées pour la première fois. En revanche, pour ce qui est des constructions figées, « se poser le problème de l'origine d'une séquence donnée implique que la structure n'est la création libre et régulière d'un locuteur mais que la combinaison lui est imposée et que cet agencement a une source historique, même si elle ne nous est plus accessible. Il est donc naturel qu'on se pose le problème de l'origine des séquences figées. Cette question a fait l'objet de recherches universitaires (P. Guiraud 1980 ; R. Martin 1996) ainsi que de nombreuses publications pour le grand public (Rey et Chanterreau 1979 ; Cl. Duneton 1990) ».

Sous cette optique et pour ce qui est des noms composés de structure $N(E + DET:G)N:G$, nous observons que le figement peut avoir une origine « externe » et faire référence à :

⁴⁶ e-volos.blogspot.com/2008/06/blog-post_19.html

⁴⁷ www.greeklaws.com/fora/lofiversion/index.php?t645.html

⁴⁸ pitsirikos.blogspot.com/2006/01/blog-post_17.html

⁴⁹ www.simerini.com.cy/nqcontent.cfm?a_id=203762

⁵⁰ www.tovima.gr/print_article.php?e=B&f=14843&m=S01&aa=2

- des événements historiques :

δέκα πληγές του Φαραώ

dix plaies d’Egypte

έξοδος του Μεσολογγίου

(la Sortie de Missolonghi)

κάθοδος των μυρίων

(descente des dix mille)

- des événements mythologiques :

μήλο της έριδος

(pomme de discorde)

κουτί της Πανδώρας

(boîte de Pandore)

κέρας της Αμαλθείας

(corne d’Amalthée)

- des événements religieux :

Γη της Επαγγελίας

(pays de cocagne + la Terre promise)

(deux entrées : nom composé entièrement
figé et entité nommée de lieu)

επιφοίτηση του Αγίου Πνεύματος

(descente du Saint-Esprit)

(deux entrées : nom composé entièrement
figé et entité nommée de fait)

εβδομάδα των παθών

(semaine de la Passion)

(deux entrées : nom composé entièrement
figé et entité nommée de date)

κατακλυσμός του Νώε

(déluge de Noé)

(deux entrées : nom composé entièrement
figé et entité nommée de fait)

- la réalité sociologique :

σύνδρομο της Κατοχής

(syndrome de l'Occupation)

γενιά (Ε + του) Πολυτεχνείου

(génération de l'Ecole Polytechnique)

θέατρο του καραγκιόζη

(théâtre d'ombres)

- la tradition populaire :

γιοφύρι της Άρτας

(œuvre inachevée + pont d'Arta)

(deux entrées : nom composé entièrement

figé et entité nommée de lieu)

ζωνάρι της Παναγιάς

(arc en ciel)

Remarquons qu'un grand nombre de ces séquences peut avoir un double emploi : un emploi propre (ou canonique), qui fait référence à l'événement de leur origine, et un emploi métaphorique. Par exemple, *κάθοδος των μυρίων*/descente des dix mille peut désigner le fait historique de la descente des dix mille dans son emploi propre, mais aussi un 'déplacement de gens en masse, surtout vers les régions littorales' dans son emploi métaphorique. Pareillement, *γιοφύρι της Άρτας*/pont d'Arta peut désigner le pont qui a été construit à Arta dans son emploi propre ou bien 'une œuvre qui ne s'achève jamais' dans son emploi métaphorique. Enfin, *εβδομάδα των παθών*/semaine sainte peut désigner la semaine sainte dans son emploi propre mais aussi une 'période très difficile' dans son emploi métaphorique etc.

Notons, cependant, que le critère de l'étymologie se révèle peu opératoire afin de distinguer les noms composés de structure $N (E + DET:G) N:G$ des $N (E + DET:G) N:G$ libres, car il repose sur des analyses qui dépassent le cadre méthodologique de la présente étude (cf. aussi S. Voyatzi 2006 : 75).

La notion de figement peut s'appliquer à toutes les catégories grammaticales comme les noms, les adjectifs, les déterminants, les adverbes etc. Nous avons jusqu'à présent examiné les critères généraux et les propriétés générales du figement, tout en les appliquant à la structure spécifique des $N (E + DET:G) N:G$ (cf. I, 2.2.1-2.2.11). « C'est dans le cadre des différentes catégories grammaticales que peut se faire l'analyse du phénomène de figement

avec la précision voulue » (G. Gross 1996a : 9). Pour cette raison, nous compléterons l'étude de la notion de figement au sein des noms composés $N(E + DET:G) N:G$, en examinant leur structure lexicale interne (cf. III, 6.1-6.8).

2.3 Distinction des $N(E + DET:G) N:G$ par rapport au degré de leur figement

Nous avons appliqué les critères présentés dans I, 2.2.1-2.2.9 aux $N(E + DET:G) N:G$ que nous avons recensés (cf. II, 4.2). Les $N(E + DET:G) N:G$ recensés présentent des degrés divers de figement. Pour des raisons de terminologie et de traitement automatique, nous les avons divisés en deux catégories : les $N(E + DET:G) N:G$ entièrement figés et les $N(E + DET:G) N:G$ semi-figés (ou productifs) (cf. aussi III, 6.4 et 6.5).

2.3.1 Les $N(E + DET:G) N:G$ entièrement figés

Les $N(E + DET:G) N:G$ entièrement figés sont des séquences qui ne permettent aucune variation d'ordre, notamment, syntaxique et lexical. A titre indicatif, citons l'exemple :

σανίδα σωτηρίας

planche salut-*Gfs*

(planche de salut)

Nous observons que toute sorte de variation y est interdite :

(i) au niveau syntaxique :

- toute détermination devant le deuxième composant nominal est interdite :

**σανίδα (της + μιας) σωτηρίας*

*planche (le-*Gfs* + un-*Gfs*) salut-*Gfs*

(*planche (du + d'un) salut)

- l'insertion de modificateurs adjectivaux n'est pas autorisée :

**σανίδα απρόσμενης σωτηρίας*

*planche inattendu-*Gfs* salut-*Gfs*

(*planche de salut inattendu)

- la permutation des composants n'est pas permise :

**σωτηρίας σανίδα*

**salut-Gfs planche*

(**de salut planche*)

- la réduction du deuxième composant nominal n'est pas autorisée non plus :

σανίδα ≠ σανίδα σωτηρίας

planche ≠ planche salut-Gfs

(*planche ≠ planche de salut*)

- (ii) au niveau lexical, la substitution des composants nominaux par d'autres substantifs nominaux sémantiquement voisins est inacceptable :

**πλάκα σωτηρίας*

(*N parasynonyme*)

**plaque salut-Gfs*

(**plaque de salut*)

**σανίδα διάσωσης*

(*N:G synonyme*)

**planche sauvetage-Gfs*

(**planche de sauvetage*)

Dans le cas de la séquence *σανίδα σωτηρίας*/(traduction), nous avons affaire à une séquence entièrement figée. Même du point de vue sémantique, le sens de cette suite n'est pas compositionnel mais figé.

A noter qu'au sein des séquences figées nous pouvons retrouver des noms composés communs, des entités nommées (cf. N. Chinchor 1998), des termes de spécialité (cf. T. Cabré 1998) et des syntagmes nominaux figés qui font partie d'une phrase simple figée et qui ont un emploi autonome en dehors de l'expression figée (cf. III, 6.4-6.8).

A titre indicatif, citons les exemples suivants :

- noms composés communs :

έξοδος κινδύνου

(*issue de secours*)

κέρας της Αμαλθείας

(corne d'Amalthée)

μερίδα του λέοντος

(part du lion)

- entités nommées :

Γη της Επαγγελίας

(entité nommée de « lieu »)

(Terre promise)

Αφροδίτη της Μήλου

(entité nommée d'« œuvre »)

(Vénus de Milo)

γέφυρα των Στεναγμών

(entité nommée de « lieu »)

(pont des Soupirs)

- termes de spécialité :

δέντρο της ζωής

(arbre de la vie) (domaine : Anatomie)

κύκλος του οξυγόνου

(cycle de l'oxygène) (domaine : Biologie)

- syntagmes nominaux figés qui font partie d'une phrase simple figée et qui ont un emploi autonome en dehors de l'expression figée :

λίθος του αναθέματος

(pierre de l'anathème)

ασκοί του Αιόλου

(outre d'Eole)

Parmi les séquences recensées, les $N (E + DET:G) N:G$ entièrement figés ne sont pas les plus fréquents. Nous avons repéré des structures qui donnent lieu à plusieurs variations et d'autres qui donnent lieu à un paradigme lexical productif. Nous allons détailler ces structures dans III, 6.

2.3.2 Les $N(E + DET:G)N:G$ semi-figés (ou productifs)

En ce qui concerne les $N(E + DET:G)N:G$ semi-figés (cf. aussi III, 6.5), nous pouvons les subdiviser en fonction de la nature de leur variabilité, en deux catégories :

- 1) les $N(E + DET:G)N:G$ semi-figés permettant un nombre de variantes restreint et, éventuellement, calculable.

Cette catégorie réunit les $N(E + DET:G)N:G$ semi-figés dont les variantes se situent essentiellement au niveau lexical et affectent les composants nominaux. A titre d'illustration, citons les exemples suivants :

- paradigme restreint du nom-tête :

(τιμή + αξία) αγοράς
(prix + valeur) d'achat

- paradigme restreint du deuxième composant nominal :

γρίπη των (πτηνών + πουλερικών)
grippe les-*Gnp* (oiseaux-*Gnp* + volailles-*Gnp*)
(grippe aviaire)

- nombre du deuxième composant nominal :

διάδρομος (προσγείωσης + προσγειώσεων)
piste (atterrissage-*Gfs* + atterrissages-*Gfs*)
(piste d'atterrissage)

- 2) les $N(E + DET:G)N:G$ semi-figés dont le deuxième composant nominal est libre.

Cette catégorie réunit les $N(E + DET:G)N:G$ semi-figés dont le composant nominal au génitif donne lieu à un paradigme ouvert. A titre d'illustration citons l'exemple suivant :

γραφείο (αλλοδαπών + πληροφοριών + δευτεροβάθμιας εκπαίδευσης + δημοσίων σχέσεων + διανομής + εισαγωγής + εργασίας + μελετών + παραγωγής + προώθησης + στοιχημάτων + τύπου + απολεσθέντων + περιβάλλοντος + σταδιοδρομίας + ...)
(bureau (d'étrangers + de renseignements + d'enseignement secondaire + de relations publiques + de distribution + d'importation + de travail + d'études + de production +

de promotion + de paris + de presse + d'objets perdus + d'environnement + de carrière + ...))

Nous avons choisi d'inclure ces cas de figure dans notre étude, notamment, à cause de l'absence systématique de déterminant. Nous adoptons la position de G. Gross (1990), qui souligne que « le plus petit degré de figement suffit pour considérer un syntagme comme figé ».

Pour conclure, signalons que la notion de nom composé, qui régit cette étude, englobe des termes allant des séquences entièrement figées jusqu'à des séquences qui présentent des centaines de réalisations possibles. Dans le cadre de cette étude, le terme « nom composé » inclut les $N(E + DET:G) N:G$ entièrement figés, les syntagmes nominaux semi-figés (ou productifs), les entités nommées, les termes de spécialité et les syntagmes nominaux figés qui font partie d'une phrase simple figée et qui ont un emploi autonome en dehors de l'expression figée.

2.3.3 Conclusion

Dans ce chapitre, nous avons abordé le phénomène du figement dans le but d'effectuer la distinction entre séquences figées de structure $N(E + DET:G) N:G$ et séquences libres de même structure. Nous avons présenté les critères proposés par G. Gross (1996a : 9-22) que nous avons appliqués à nos données lexicales afin de qualifier une structure $N(E + DET:G) N:G$ de nom composé (ou séquence nominale figée) et de justifier sa représentation dans nos ressources lexicales. Nous avons constaté que la distinction entre structures libres et structures figées ne pourrait pas être fondée uniquement sur des critères sémantiques, évoquant le concept traditionnel de la « non-compositionnalité du sens ». Les critères sur lesquels nous nous sommes appuyée mettent en évidence les contraintes syntaxiques et lexicales des $N(E + DET:G) N:G$, tout en démontrant que les aspects du phénomène de figement sont multiples et variés. Les exemples que nous avons exposés, tout au long de ce chapitre, illustrent la dimension scalaire du phénomène de figement au sein des séquences nominales étudiées ici. Nous constatons donc qu'il existe un continuum entre les $N(E + DET:G) N:G$ entièrement figés et les $N(E + DET:G) N:G$ libres.

DEUXIEME PARTIE : LE RECENSEMENT DES DONNEES

Chapitre 3. Brève description du dictionnaire électronique morphologique

3.1 Introduction

Rappelons que les noms composés posent problème lors de l'analyse lexicale des textes, qui constitue l'étape préalable à toute analyse ou génération ou encore traduction automatiques des textes. En d'autres termes, si un nom composé n'est pas reconnu en tant qu'une seule unité lexicale, il ne pourra pas être analysé correctement et, par conséquent, ne sera pas proprement interprété ou traduit. Dans le cadre de ce travail, nous avons adopté une méthode linguistique (ou symbolique) pour l'étude des noms composés. De façon générale, pour la reconnaissance des mots composés les systèmes symboliques s'articulent autour de deux possibilités : les règles générales (exprimées sous forme de graphes, expressions régulières ou algorithmes) ou les dictionnaires électroniques. La construction des dictionnaires électroniques exige le recensement – le plus exhaustif possible – de noms composés. Dans le cadre de cette étude, nous avons opté pour la construction manuelle de ressources lexicales et nous avons donc effectué le recensement de noms composés sur la base de critères de délimitation (cf. I, 2.2.1-2.2.9).

L'étape du recensement des noms composés met en évidence les problèmes qui touchent, d'une part, à la définition et délimitation des noms composés ainsi qu'aux méthodes de leur représentation formalisée et, d'autre part, à la maintenance et la mise à jour des ressources construites. Rappelons que, dans le cadre de la présente étude, nous avons étudié les propriétés spécifiques des $N(E + DET:G)N:G$ (cf. I, 2.3.1-2.3.2 et III, 6.2-6.8). Nous en avons repéré des sous-catégories qui pourraient s'avérer opérationnelles en vue de l'analyse automatique des textes (cf. III, 6.2-6.8). Ceci nous a permis, tout d'abord, d'améliorer la présentation des données dans le dictionnaire électronique des noms composés du grec moderne existant (cf. T. Kyriacopoulou *et al.* 2002). Ensuite, le repérage en sous-catégories nous a permis une nouvelle proposition pour l'organisation et la structuration des données.

Dans ce chapitre, nous présentons brièvement le dictionnaire électronique morphologique existant du grec moderne.

3.2 Le dictionnaire électronique morphologique des noms composés du grec moderne

Rappelons que dans le cadre de la construction du dictionnaire électronique morphologique des noms composés, la définition du mot composé est motivée par la perspective du traitement automatique des données (M. Silberztein 1990) ; elle est donc purement formelle et dépendante de la définition du mot simple (cf. I, 1.1.1) :

- un mot simple est défini formellement comme une suite de caractères comprise entre deux délimiteurs consécutifs (*i.e.* blanc, tiret, apostrophe, ponctuation). Les mots composés soudés (e.g. *αγριολούλουδο*/fleur des champs)⁵¹ sont considérés comme des mots simples (cf. I, 1.1.1) et font partie du dictionnaire électronique morphologique des mots simples (DELAG⁵²). Le DELASG⁵³ comprend actuellement 251 574 entrées lexicales (ou lemmes), qui correspondent à 2 029 797 formes fléchies. Ces entrées se présentent sous la forme canonique, accompagnés de leur catégorie grammaticale et d'un code flexionnel qui permet à un programme de flexion d'en générer automatiquement les formes fléchies. A noter que d'autres informations peuvent figurer dans le DELASG (e.g. traits sémantiques, le trait « SV » pour les formes de la langue savante) ;
- un mot composé se définit formellement comme une suite incluant au moins deux mots simples et au moins un séparateur. Selon la partie du discours, les mots composés, tout comme les mots simples, peuvent être répartis dans toutes les catégories grammaticales, notamment en :

noms :

φούρνος μικροκυμάτων/four à micro-ondes

adverbes :

το μήνα που δεν έχει Σάββατο/pendant la semaine des quatre jeudis

conjonctions (de subordination) :

αν και/bien que

⁵¹ Cependant, lorsque les noms composés $N(E + DET:G)N:G$ présentent une variante soudée (e.g. *βόμβα υδρογόνου* = *υδρογονοβόμβα*/bombe à hydrogène, nous effectuons la mise en relation des deux variantes dans les tables du lexique-grammaire (cf. V, 11.2.1 et 11.3.3)).

⁵² Dictionnaire électronique morphologique des formes fléchies des mots simples.

⁵³ Dictionnaire électronique morphologique des formes canoniques des mots simples.

verbes :

παίρνω στα σοβαρά/prendre au sérieux

prépositions :

δίπλα σε/à côté de

Le DELACG⁵⁴ comporte actuellement 36 701 formes au total, qui correspondent aux 161 145 formes fléchies du DELACFG⁵⁵. Comme le DELASG, les entrées du DELACG sont sous leur forme canonique, accompagnés de leur catégorie grammaticale et d'un code flexionnel (et éventuellement d'autres informations, comme par exemple des filtres concernant des restrictions de flexion, les traits sémantiques ou le trait « SV »). La flexion des mots composés prend appui sur celle des noms simples. Les formes fléchies des mots composés sont générées automatiquement à partir de la forme canonique à l'aide d'un programme de flexion (cf. V, 11.3.1.3).

Notons que les noms composés comprennent, d'une part, les noms composés binaires, c'est-à-dire, les noms composés de deux mots « pleins » et, d'autre part, les noms composés complexes (ou surcomposés), qui comportent les expansions des noms composés binaires. Les noms composés binaires et les noms composés complexes (ou surcomposés) sont traités de la même manière dans le dictionnaire électronique morphologique. Nous donnons ci-après des exemples de cette classification :

⁵⁴ Dictionnaire électronique morphologique des formes canoniques des mots composés.

⁵⁵ Dictionnaire électronique morphologique des formes fléchies des mots composés.

i) classes élémentaires : noms composés binaires

Structures morphosyntaxiques		Exemples
<i>NN</i>	<i>Nom Nom</i>	<p>παιδί θαύμα (enfant prodige)</p> <p>ταξίδι αστραπή (voyage éclair)</p>
<i>AN</i>	<i>Adjectif Nom</i>	<p>παιδική χαρά (aire de jeux)</p> <p>φαύλος κύκλος (cercle vicieux)</p> <p>γενική συνέλευση (assemblée générale)</p>
<i>N (E + DET:G) N:G</i>	<i>Nom (E + Déterminant au génitif) Nom au génitif</i>	<p>σύνοδος κορυφής (sommet)</p> <p>ανιχνευτής ψεύδους (détecteur de mensonges)</p> <p>μερίδα του λέοντος (part du lion)</p> <p>μαλλί της γριάς (barbe à papa)</p> <p>πιάτο (E + της) ημέρας (plat du jour)</p>
<i>N PrépN</i>	<i>Nom Préposition Nom</i>	<p>εισιτήριο μετ' επιστροφής (billet aller retour)</p> <p>άδεια άνευ αποδοχών (congé non payés)</p>

ii) expansions des AN et N (*E + DET:G*) *N:G* : noms composés complexes (ou surcomposés)

Structures morphosyntaxiques		Exemple
<i>AAN</i>	<i>Adjectif Adjectif Nom</i>	<i>ακαθάριστο εθνικό προϊόν</i> (produit national brut) <i>αεριούχο μεταλλικό νερό</i> (eau minérale gazeuse)
<i>NA:G N:G</i>	<i>Nom Adjectif au génitif Nom au génitif</i>	<i>ώρες κοινής ησυχίας</i> (heures de tranquillité commune) (tr. littérale) <i>φόρος προστιθέμενης αξίας</i> (taxe sur la valeur ajoutée)
<i>A N (E + DET:G) N:G</i>	<i>Adjectif Nom (E + Déterminant au génitif) Nom au génitif</i>	<i>ολική έκλειψη σελήνης</i> (éclipse lunaire totale) <i>πρόδρομα φαινόμενα νόσου</i> (phénomènes précurseurs de maladie) (tr. littérale) <i>τελευταίος τροχός της αμάξης</i> (cinquième roue du carrosse)
<i>ANA:G N:G</i>	<i>Adjectif Nom Adjectif au génitif Nom au génitif</i>	<i>νομικό πρόσωπο δημοσίου δικαίου</i> (personne morale de droit public)
<i>N DET:G A:G N:G</i>	<i>Nom Déterminant au génitif Adjectif au génitif Nom au génitif</i>	<i>χώρα της φαιδράς πορτοκαλέας</i> (pays de l'oranger joyeux) (tr. littérale)
<i>N DET:G Adv A:G N:G</i>	<i>Nom Déterminant au génitif Adverbe Adjectif au génitif Nom au génitif</i>	<i>ρήτρα του μάλλον ευνοούμενου κράτους</i> (clause de la nation la plus favorisée)
<i>N Prép A:A N:A</i>	<i>Nom Préposition Adjectif à l'accusatif Nom à l'accusatif</i>	<i>άτομο με ειδικές ανάγκες</i> (personne à mobilité réduite)

Chapitre 4. Le recensement

4.1 Introduction

Selon M. Silberztein (1993b: 405), le recensement – manuel ou automatique – des noms composés s'avère une tâche difficile, notamment à cause du problème de la délimitation entre noms composés (entrées lexicales), noms semi-figés (séquences productives) et syntagmes nominaux libres (séquences compositionnelles), problème déjà largement discuté par de nombreux lexicographes et linguistes.

Pour effectuer le recensement manuel des mots composés, les linguistes disposent d'un nombre important de sources documentaires : des dictionnaires et des grammaires de la langue générale, des lexiques de langues de spécialité, des corpus de textes écrits sur support papier ou numériques et le Web. Cependant, en raison de la créativité lexicale et des problèmes de délimitation déjà discutés (cf. I, 3.2.1-3.2.10 et I, 3.3.1-3.3.2), il est impossible de dresser une liste exhaustive des noms composés en reposant uniquement sur ces ressources.

Dans l'intention d'inventorier les noms composés $N(E + DET:G)N:G$ appartenant à l'usage courant de la langue grecque, nous avons décidé de procéder en trois étapes. Dans un premier temps, nous parlerons de la validation manuelle des entrées $N(E + DET:G)N:G$, figurant dans le dictionnaire électronique des mots composés (DELAC) existant. Dans un deuxième temps, nous avons compilé les dictionnaires édités et les grammaires d'usage du grec moderne. Enfin, nous avons complété cette liste de façon semi-automatique grâce au dépouillement d'un corpus de presse numérique. Nous nous sommes servie du Web pour la validation des exemples douteux.

Notons, enfin, que dans ce chapitre nous présentons tous les différents types de séquences que nous avons rassemblées à l'aide des méthodes de recensement précitées. Une fois les données rassemblées à partir des diverses sources, nous nous sommes intéressée à regrouper les $N(E + DET:G)N:G$ selon leur structure interne (cf. II, 4.2.4). Les choix effectués à partir de nos données sont présentés en détail dans III, 6.2-6.8 et V, 11.3.1-11.3.3.

4.2 Méthodes de recensement

4.2.1 Validation des entrées $N(E + DET:G) N:G$ du DELAC existant

La première étape du recensement a été consacrée à la validation des noms composés de structure $N(E + DET:G) N:G$ du DELAC actuel. Nous avons appelé DELAC-NNG le sous-lexique du DELAC qui rassemble les $N(E + DET:G) N:G$.

Dans l'état actuel, le DELAC-NNG décrit la morphologie et la flexion de 9 358 noms composés (binaires et complexes). Par convention, les noms composés $N(E + DET:G) N:G$ sont entrés dans le dictionnaire accompagnés des informations suivantes (cf. aussi V, 11.3.1.1) :

- i) des codes assignés à chacun des éléments simples qui le composent (N265, A10, DET etc.) ;
- ii) du code « N » signalant que l'ensemble de l'entrée constitue un nom ;
- iii) du signe « " » quand il y a déplacement de l'accent lors de la flexion ;
- iv) des filtres pour la gestion des contraintes flexionnelles (e.g. « S » quand le nom composé est fléchi uniquement au singulier, « -GS2 » ou « -GP2 » pour désigner que le deuxième composant nominal reste invariable respectivement au génitif du singulier ou au génitif du pluriel.

A titre d'illustration, citons ci-dessous un petit extrait du DELAC-NNG avant sa validation⁵⁶ :

```
εξέτα"ση.N265 αίματος.N363,N,-GS2
εξέτα"ση.N265 αίματος.N363,N,-GS2
εξέτα"ση.N265 μαρτύ"ρων.N186,N,S,-GP2
εξέτα"ση.N265 μαρτύ"ρων.N43300,N,S,-GP2
εξέτα"ση.N265 οπτικού.A10 πεδίου.N311,N,-GS2,-GS3
εξέτα"ση.N265 των.DET μαρτύ"ρων.N186,N,-GP3
εξίσω"ση.N265 διάχυ"σης.N265,N,-GS2
εξίσω"ση.N265 ηλικίας.N232 συνταξιοδότη"σης.N265,N,S,-GS2,-GS3
εξαγγελία.N232 υποψηφιότη"τας.N236,N,-GS2
εξαργύρω"ση.N265 λαχείου.N311,N,-GS2
εξαργύρω"ση.N265 λαχείων.N311,N,-GP2
εξαργύρω"ση.N265 συναλλαγματικής.N251,N,-GS2
εξαργύρω"ση.N265 συναλλαγματικών.N251,N,-GP2
εξελικτική.A10 ιστορία.N232 των.DET ειδών.N344,N,S,-GP4
εξοικονόμη"ση.N265 δυνά"μεων.N265,N,S,-GP2
εξοικονόμη"ση.N265 ενέργειας.N232,N,S,-GS2
εξοικονόμη"ση.N265 χρόνου.N121,N,S,-GS2
εξομάλυ"νση.N265 καμπύλης.N253,N,S,-GS2
```

⁵⁶ Pour l'explicitation des différents codes utilisés dans le cadre du DELAC cf. V, 11.3.1.1-11.3.1.3.

εξοπλισμός.N122 γραφείου.N311,N,S,-GS2
 εξοπλισμός.N122 ζύγι"σης.N265,N,S,-GS2
 εξοφλητική.A10 απόδει"ξη.N265 εκκαθάρι"σης.N265 αποδοχών.N251,N,-GS3,-GP4
 εξυγία"νση.N265 εδά"φους.N356,N,-G2
 εξυπηρέτη"ση.N265 πελατών.N194,N,S,-GP2
 εξωτερική.A10 αξία.N232 του.DET χρήμα"τος.N363,N,S,-GS4

Les formes fléchies ont été générées automatiquement à partir des formes canoniques à l'aide d'un programme de flexion automatique (cf. V, 11.3.1.3). Ci-dessous l'extrait du DELACF-NNG qui a été généré pour les entrées susmentionnées :

εξέταση αίματος,.N:Nfs:Afs:Vfs
 εξέτασης αίματος, εξέταση αίματος.N:Gfs
 εξετάσεως αίματος, εξέταση αίματος.N:Gfs
 εξετάσεις αίματος, εξέταση αίματος.N:Nfp:Afp:Vfp
 εξετάσεων αίματος, εξέταση αίματος.N:Gfp
 εξέταση αίματος,.N:Nfs:Afs:Vfs
 εξέτασης αίματος, εξέταση αίματος.N:Gfs
 εξετάσεως αίματος, εξέταση αίματος.N:Gfs
 εξετάσεις αίματος, εξέταση αίματος.N:Nfp:Afp:Vfp
 εξετάσεων αίματος, εξέταση αίματος.N:Gfp
 εξέταση μαρτύρων,.N:Nfs:Afs:Vfs
 εξέτασης μαρτύρων, εξέταση μαρτύρων.N:Gfs
 εξετάσεως μαρτύρων, εξέταση μαρτύρων.N:Gfs
 εξέταση μαρτύρων,.N:Nfs:Afs:Vfs
 εξέτασης μαρτύρων, εξέταση μαρτύρων.N:Gfs
 εξετάσεως μαρτύρων, εξέταση μαρτύρων.N:Gfs
 εξέταση οπτικού πεδίου,.N:Nfs:Afs:Vfs
 εξέτασης οπτικού πεδίου, εξέταση οπτικού πεδίου.N:Gfs
 εξετάσεως οπτικού πεδίου, εξέταση οπτικού πεδίου.N:Gfs
 εξετάσεις οπτικού πεδίου, εξέταση οπτικού πεδίου.N:Nfp:Afp:Vfp
 εξετάσεων οπτικού πεδίου, εξέταση οπτικού πεδίου.N:Gfp
 εξέταση των μαρτύρων,.N:Nfs:Afs:Vfs
 εξέτασης των μαρτύρων, εξέταση μαρτύρων.N:Gfs
 εξίσωση διάχυσης,.N:Nfs:Afs:Vfs
 εξίσωση διαχύσεως, εξίσωση διάχυσης.N:Nfs:Afs:Vfs
 εξίσωσης διάχυσης, εξίσωση διάχυσης.N:Gfs
 εξισώσεως διάχυσεως, εξίσωση διάχυσης.N:Gfs
 εξισώσεις διάχυσης, εξίσωση διάχυσης.N:Nfp:Afp:Vfp
 εξισώσεως διάχυσεως, εξίσωση διάχυσης.N:Nfp:Afp:Vfp
 εξισώσεων διάχυσης, εξίσωση διάχυσης.N:Gfp
 εξισώσεων διάχυσεως, εξίσωση διάχυσης.N:Gfp
 εξίσωση ηλικίας συνταξιοδότησης,.N:Nfs:Afs:Vfs
 εξίσωσης ηλικίας συνταξιοδότησης, εξίσωση ηλικίας συνταξιοδότησης.N:Gfs
 εξαγγελία υποψηφιότητας,.N:Nfs:Afs:Vfs
 εξαγγελία υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Nfs:Afs:Vfs
 εξαγγελίας υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Gfs
 εξαγγελίας υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Gfs
 εξαγγελίες υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Nfp:Afp:Vfp
 εξαγγελίες υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Nfp:Afp:Vfp
 εξαγγελιών υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Gfp
 εξαγγελιών υποψηφιότητος, εξαγγελία υποψηφιότητας.N:Gfp
 εξαργύρωση λαχείου,.N:Nfs:Afs:Vfs
 εξαργύρωσης λαχείου, εξαργύρωση λαχείου.N:Gfs
 εξαργύρωση λαχείων, εξαργύρωση λαχείου.N:Nfs:Afs:Vfs

εξαργύρωση λαχείων, εξαργύρωση λαχείου.N:Gfs
 εξαργύρωση συναλλαγματικής, .N:Nfs:Afs:Vfs
 εξαργύρωση συναλλαγματικής, εξαργύρωση συναλλαγματικής.N:Gfs
 εξαργύρωση συναλλαγματικών, εξαργύρωση συναλλαγματικής.N:Nfs:Afs:Vfs
 εξαργύρωση συναλλαγματικών, εξαργύρωση συναλλαγματικής.N:Gfs
 εξελικτική ιστορία των ειδών, .N:Nfs:Afs:Vfs
 εξελικτικής ιστορίας των ειδών, εξελικτική ιστορία των ειδών.N:Gfs
 εξοικονόμηση ενέργειας, .N:Nfs:Afs:Vfs
 εξοικονόμησης ενέργειας, εξοικονόμηση ενέργειας.N:Gfs
 εξοικονομήσεως ενέργειας, εξοικονόμηση ενέργειας.N:Gfs
 εξοικονόμηση δυνάμεων, .N:Nfs:Afs:Vfs
 εξοικονόμησης δυνάμεων, εξοικονόμηση δυνάμεων.N:Gfs
 εξοικονόμηση χρόνου, .N:Nfs:Afs:Vfs
 εξοικονόμησης χρόνου, εξοικονόμηση χρόνου.N:Gfs
 εξομάλυνση καμπύλης, .N:Nfs:Afs:Vfs
 εξομάλυνσης καμπύλης, εξομάλυνση καμπύλης.N:Gfs
 εξοπλισμός γραφείου, .N:Nms
 εξοπλισμού γραφείου, εξοπλισμός γραφείου.N:Gms
 εξοπλισμό γραφείου, εξοπλισμός γραφείου.N:Ams
 εξοπλισμέ γραφείου, εξοπλισμός γραφείου.N:Vms
 εξοπλισμός ζύγισης, .N:Nms
 εξοπλισμού ζύγισης, εξοπλισμός ζύγισης.N:Gms
 εξοπλισμό ζύγισης, εξοπλισμός ζύγισης.N:Ams
 εξοπλισμέ ζύγισης, εξοπλισμός ζύγισης.N:Vms
 εξοφλητική απόδειξη εκκαθάρισης αποδοχών, .N:Nfs:Afs:Vfs
 εξοφλητικής απόδειξης εκκαθάρισης αποδοχών, εξοφλητική απόδειξη
 εκκαθάρισης αποδοχών.N:Gfs
 εξοφλητικές αποδείξεις εκκαθάρισης αποδοχών, εξοφλητική απόδειξη
 εκκαθάρισης αποδοχών.N:Nfp:Afp:Vfp
 εξοφλητικών αποδείξεων εκκαθάρισης αποδοχών, εξοφλητική απόδειξη
 εκκαθάρισης αποδοχών.N:Gfp
 εξυγίανση εδάφους, .N:Nfs:Afs:Vfs
 εξυγίανσης εδάφους, εξυγίανση εδάφους.N:Gfs
 εξυγιάνσεως εδάφους, εξυγίανση εδάφους.N:Gfs
 εξυγιάνσεις εδαφών, εξυγίανση εδάφους.N:Nfp:Afp:Vfp
 εξυγιάνσεως εδαφών, εξυγίανση εδάφους.N:Gfp
 εξυπηρέτηση πελατών, .N:Nfs:Afs:Vfs
 εξυπηρέτησης πελατών, εξυπηρέτηση πελατών.N:Gfs
 εξυπηρετήσεως πελατών, εξυπηρέτηση πελατών.N:Gfs
 εξωτερική αξία του χρήματος, .N:Nfs:Afs:Vfs
 εξωτερικής αξίας του χρήματος, εξωτερική αξία του χρήματος.N:Gfs

La validation comprend les tâches suivantes : vérification du statut de nom composé grâce à l'aide des critères formels de délimitation des noms composés (cf. I, 2.2.1-2.2.11), élimination des doublons, correction et homogénéisation des codes flexionnels et des filtres morphologiques assignés aux entrées.

Nous exposerons par la suite les problèmes auxquels nous avons fait face lors de la validation du DELAC-NNG.

Tout d'abord, il n'existe pas – à notre connaissance – suffisamment de documentation disponible concernant la constitution du DELAC en grec moderne (au moins pour ce qui est

des entrées $N(E + DET:G)N:G$. Dans T. Kyriacopoulou *et al.* (2002) les auteurs s'intéressent à la flexion des noms composés, mais pas aux sources documentaires qui ont été consultées ni aux critères de délimitation des noms composés.

De plus, l'organisation des entrées du DELAC-NNG ne tient pas compte des différentes sous-catégories des $N(E + DET:G)N:G$ (I, 2.3.1-2.3.2, II, 4.2.4 et III, 6.2-6.8), ce qui complique l'exploitation des données pour des éventuelles applications spécifiques. Nous y retrouvons donc toutes les sous-catégories confondues, à savoir :

- des noms composés entièrement figés :

μαλλιά αγγέλου

(cheveux d'ange) ;

- des syntagmes nominaux semi-figés (ou productifs) :

μονάδα (αεροπορίας + απεξάρτησης + αφαλάτωσης + βιολογικού καθαρισμού + ελέγχου + εντατικής θεραπείας + εντατικής παρακολούθησης + παστερίωσης + υγειονομικής περίθαλψης + υπερήχου + παραγωγής +...)

(unité (d'aviation + de désintoxication + de dessalage + d'épuration biologique + de contrôle + de soins intensifs + de suivi intensif + de pasteurisation + de soins sanitaires + d'ultrason + de production +...)) ;

- des entités nommées :

Βήμα της Κυριακής

(*Vima tis Kyriakis*) (entité nommée d'« œuvre »)

Γιατροί του Κόσμου

(Médecins du Monde) (entité nommée d'« organisation ») ;

- des termes de spécialité :

εύρος ζώνης

(largeur de bande) (domaine : Electronique)

λόβιο παρεγκεφαλίδας

(lobule du cervelet) (domaine : Anatomie)

- des syntagmes nominaux figés faisant partie de phrases simples figées et qui n'ont pas d'emploi autonome en dehors de l'expression figée :

στήλη άλατος

(bouche bée)

κώδων του κινδύνου

(tocsin)

- des noms prédicatifs :

(κάνω) κίνηση αντιπερισπασμού

((faire) (acte de diversion))

Visiblement, le dictionnaire électronique morphologique constitue la méthode de représentation formalisée qui a été choisie au préalable pour l'ensemble des noms « multi-mots » qui présentent un degré de figement quelles que soient leurs caractéristiques spécifiques. Cependant, ces caractéristiques (cf. III, 6.2-6.8) permettent de repérer plusieurs sous-catégories au sein des $N(E + DET:G)N:G$, qui ont des spécificités et qui sont distinctes aussi bien du point de vue linguistique que du point de vue du TAL. Dans ce but, nous proposons un codage permettant d'identifier la sous-catégorie des entrées et différentes méthodes de représentation formalisée adaptées aux différentes sous-catégories identifiées (cf. V, 11.3.1-11.3.3).

Par ailleurs, nous avons observé que nombre de séquences libres (ou compositionnelles) font partie des entrées du DELAC-NNG. Par exemple :

λειτουργία σχολείου

(opération d'école)

Pourquoi aurait-on décidé d'inclure uniquement *λειτουργία σχολείου*/opération d'école et non pas *λειτουργία πανεπιστημίου*/opération d'université, *λειτουργία νηπιαγωγείου*/opération de crèche, *λειτουργία εργοστασίου*/opération d'usine, *λειτουργία τράπεζας*/opération de banque etc. ?

Nous proposons l'élimination des séquences libres figurant dans le DELAC-NNG suite à l'application des critères de délimitation (cf. I, 3.2.1-3.2.10).

Dans certains cas, nous n'avons pas pu décider sur le statut des entrées, car, faute de documentation pertinente, il n'était pas possible de comprendre s'il s'agissait d'une séquence libre ou d'un terme de spécialité. A titre indicatif citons les exemples suivants :

ανακάτεμα (σούπας + υλικών + των ρούχων...)

(mélange (de la soupe + des ingrédients + des vêtements...))

αποστολή (επιστολής + εμπορευμάτων +...)

(envoi (de lettre + de marchandises + ...))

Nous avons également noté que, dans le DELAC-NNG, les différentes variantes (orthographiques, lexicales, syntaxiques) d'un même nom composé $N (E + DET:G) N:G$ sont représentées en tant qu'entrées distinctes et ne sont pas reliées entre elles⁵⁷. Par exemple, pour le nom composé :

ένστικτο (E + της) μητρότητας

instinct (E + la-_{Gfs}) maternité-_{Gfs}

(instinct de maternité)

nous trouvons deux entrées distinctes dans le DELAC-NNG, l'une avec déterminant et l'autre sans déterminant :

ένστι"κτο.N301 μητρότητας.N236,N,S,-GS2

ένστι"κτο.N301 της.DET μητρότητας.N236,N,S,-GS3

Pareillement pour les variantes orthographiques du nom composé :

(αβγό + αυγό) του Κολόμβου

œuf Colomb-_{Gms}

(œuf de Colomb)

nous retrouvons dans le DELAC-NNG les deux entrées suivantes :

αβγό.N300 του.DET Κολόμβου.N121,N,S,-GS3

αυγό.N300 του.DET Κολόμβου.N121,N,S,-GS3

⁵⁷ Pour l'étude, la mise en relation et la représentation formalisée des différentes variantes cf. aussi II, 5 et V, 11.2.1 et 11.3.3.

Parfois deux variantes (*i.e.* avec et sans déterminant) d'une séquence $N (E + DET:G) N:G$ sont représentées dans le DELAC-NNG, alors qu'une d'entre elles constitue en effet un nom composé. Ainsi pour le nom composé :

άκρη του νήματος
 bout le-*Gns* fil-*Gns*
 (solution d'un problème)

nous retrouvons dans le DELAC-NNG la variante sans déterminant, mais celle-ci est, en effet, une séquence compositionnelle :

άκρη.N2512 του.DET νήματος.N363,N,S,-GS3
 άκρη.N2512 νήματος.N363,N,S,-GS2

Le grand nombre de doublons (plus de 1 000) vient s'ajouter à la liste des problèmes rencontrés lors de la validation manuelle du DELAC-NNG. Ainsi, le nom composé *ταινία μικρού μήκους*/court métrage est codé deux fois, *έγκλημα πολέμου*/crime de guerre trois fois, *ένταλμα σύλληψης*/mandat d'arrêt quatre fois, etc. :

Ταινία.N232 Μικρού.A10 Μήκους.N344,N,-GS2,-GS3
 ταινία.N232 μικρού.A10 μήκους.N344,N,-GS2,-GS3

έγκλημα.N365 πολέμου.N125,N,-GS2
 έγκλημα.N365 πολέμου.N125,N,-GS2
 έγκλημα.N365 πολέμου.N125,N,-GS2

ένταλμα.N365 σύλληψης.N265,N,-GS2
 ένταλμα.N365 σύλληψης.N265,N,-GS2
 ένταλμα.N365 σύλληψης.N265,N,-GS2
 ένταλμα.N365 σύλληψης.N265,N,-GS2

Des problèmes d'incohérence au niveau du codage sont aussi à signaler : les mêmes mots ne sont pas toujours assignés au même code flexionnel. A titre indicatif, citons les exemples suivants :

ακτίνα.N202 ακοής.N251,N,-GS2
 ακτίνα.N23203 δράσης.N261,N,-GS2

αλγόριθμος.N130 συμπίεσης.N265,N,-GS2
 αλγόριθμος.N125 διάσπασης.N265,N,-GS2

απονομή.N251 **δικαιοσύνης.N2512**, N, S, -GS2

απονομή.N251 **δικαιοσύνης.N253**, N, -GS2

Notons, cependant, que l'incohérence au niveau du codage est un problème secondaire dans la mesure où ce qui nous intéresse principalement est la génération de formes fléchies correctes : c'est le DELACFG que l'on va appliquer dans les textes pour obtenir la reconnaissance lexicale automatique. Il s'agit d'un problème qui touche à la maintenance, à la gestion et à la mise à jour du dictionnaire.

Enfin, nous avons observé que le DELAC-NNG comprend un nombre élevé d'entités nommées qui ne sont pas systématiquement validées : souvent seulement une partie de l'ensemble de l'entité nommée figure dans le dictionnaire. Nous en donnons ici quelques exemples :

Entrées du DELAC-NNG	Séquences complètes
<i>Ένωση Πτυχιούχων</i> (Union de diplômés)	<i>(Πανελλήνια Ένωση Πτυχιούχων Οινολόγων + Ένωση Πτυχιούχων Περιβαλλοντολόγων Ελλάδας + Ένωση Πτυχιούχων Τεχνικών Αεροπλάνων Ολυμπιακής Αεροπορίας + ...)</i> (Union Panhellénique de diplômés œnologues + Union de diplômés environmentalistes de Grèce + Union de diplômés techniciens d'avions d'aviation Olympique + ...) (entités nommées d'« organisation »)
<i>Ίδρυμα Μελετών</i> (Fondation d'études)	<i>(Ίδρυμα Μελετών Λαμπράκη + Ίδρυμα Μελετών Χερσονήσου του Αίμου + ...)</i> (Fondation d'études Lambraki + Fondation d'études balkaniques + ...) (entités nommées d'« organisation »)

En ce qui concerne les exemples ci-dessus, nous percevons une hésitation concernant le choix entre le codage de la forme complète de l'entité nommée ou bien de sa forme réduite, qui sert de « mot déclencheur » (D. McDonald 1996 ; N. Friburger 2002 ; C. Martineau *et al.* 2007) pour la reconnaissance (ou l'extraction) automatique des entités nommées. Ce mot déclencheur peut être interne et, dans ce cas, il fait partie de l'entité nommée :

Οργανισμός Σιδηροδρόμων Ελλάδας
(**Organisation** des chemins de fer de Grèce)

Mais, il peut également être externe et, dans ce cas, il n'appartient pas forcément à l'entité nommée et constitue son contexte d'apparition dans une occurrence donnée :

κέντρο αισθητικής Vita Plus
(**centre de beauté** Vita Plus)

Le problème de traitement des mots déclencheurs (internes ou externes) s'interpose au problème général de traitement des syntagmes nominaux semi-figés (ou productifs), qui constituent souvent eux aussi des déclencheurs d'entités nommées. Dans une perspective applicative, le recensement de tels déclencheurs s'avère très utile, puisqu'ils peuvent être utilisés pour la reconnaissance dynamique des entités nommées dans des systèmes d'analyse de requêtes ou de moteurs de recherche. Cependant, nous pensons que leur codage dans un dictionnaire électronique n'est pas la méthode de représentation la plus appropriée, notamment à cause de leur nombre fort élevé et même imprévisible. Nous proposons leur traitement à l'aide de graphes lexicalisés. En ce qui concerne les entités nommées, nous proposons d'abord la reconstitution de la forme complète et ensuite la constitution d'une liste séparée. Cette liste pourrait faire partie du dictionnaire électronique morphologique des noms composés ; dans ce cas, nous proposons l'ajout d'un code qui fournirait l'information qu'il s'agit d'une entité nommée.

L'étape de la validation manuelle du dictionnaire électronique morphologique existant s'est avérée assez coûteuse : nous y avons consacré six mois de travail. Une fois la validation terminée, nous avons estimé qu'il était utile de procéder à un recensement systématique des $N(E + DET:G) N:G$ à l'aide des méthodes décrites dans II, 4.2.2-4.2.3. Ceci nous a donné la possibilité de confronter les entrées du DELAC-NNG aux listes obtenues à l'aide de ces méthodes. Nous avons retenu environ 1 500 entrées du DELAC-NNG et nous y avons ajouté les nouveaux noms composés ainsi repérés.

4.2.2 Méthode manuelle : compilation de grammaires et de dictionnaires d'usage

Dans le but de recenser les noms composés $N(E + DET:G)N:G$, nous avons compilé manuellement :

a) les dictionnaires monolingues du grec moderne :

- de G. Babiniotis (1998) ;
- de l'Institut des Études Néohelléniques (1998) ;

b) un nombre restreint des grammaires usuelles du grec moderne :

- de M. Triantaphyllidis (2005 [1941]) ;
- d'A. Tzartanos (1989 [1946]) ;
- de D. Holton, P. Mackridge et I. Philippaki-Warbuton (2000) ;
- de C. Clairis et G. Babiniotis (2001).

Nous avons préféré ces ouvrages à leurs homologues, parce qu'étant les plus récents, ils bénéficient de données supplémentaires par rapport aux concurrents.

Grâce à cette méthode, nous avons recensé 4 200 noms composés (y compris les syntagmes nominaux semi-figés (ou productifs)).

La vaste majorité des noms composés recensés sont binaires (cf. III, 6.2), c'est-à-dire des noms composés constitués de deux noms simples « pleins ». Par exemple :

αγορά εργασίας

marché travail-Gfs

(marché du travail)

πάλη των τάξεων

lutte les-Gfp classes-Gfp

(lutte des classes)

Cependant, nous avons également inclus dans notre inventaire des formes surcomposées ternaires ou quaternaires de structure $N(E + DET:G)N:G$, c'est-à-dire des séquences dont

l'un au moins des constituants est lui-aussi un nom composé binaire. Nous retrouvons donc des formes définies par les structures morphosyntaxiques suivantes⁵⁸ :

$N (E + DET:G) N:G =: N A:Gnp N:Gnp$

αγωγός όμβριων υδάτων

collecteur pluviales-*Gnp* eaux-*Gnp*

(collecteur d'eaux pluviales)

$N (E + DET:G) N:G =: N DET:Gfs A:Gfs N:Gfs$

χάπι της επόμενης μέρας

pilule la-*Gfs* suivante-*Gfs* journée-*Gfs*

(pilule du lendemain)

$N (E + DET:G) N:G =: A N N:Gnp$

διεθνής κώδικας σημάτων

international code signaux-*Gnp*

(code international de signaux)

$N (E + DET:G) N:G =: A N N:Gfp N:Gms$

γενικός δείκτης τιμών καταναλωτή

général indice prix-*Gfp* consommateur-*Gms*

(indice général des prix à la consommation)

Signalons que, dans certains cas, les noms composés complexes (ou surcomposés) possèdent une variante réduite qui correspond à un $N (E + DET:G) N:G$ binaire⁵⁹. Par exemple :

$N (E + DET:G) N:G =: N N:Gfs (E + N:Gns)$

άδεια κυκλοφορίας (E + αυτοκινήτου)

permis circulation-*Gfs* (E + voiture-*Gns*)

(certificat d'immatriculation (E + de voiture))

Les deux variantes ont été retenues dans notre inventaire. Elles sont représentées et reliées entre elles par la même forme canonique dans les tables du lexique-grammaire (cf. II, 5.3.4 et V, 11.3.3).

⁵⁸ Les structures morphosyntaxiques des noms composés complexes (ou surcomposés) sont présentées en détail dans III, 6.3.

⁵⁹ Les variantes des noms composés complexes (ou surcomposés) sont étudiées dans II, 5.3.1-5.3.4.

L'existence des formes surcomposées met en évidence que la combinatoire syntaxique des noms composés est similaire à celle des noms simples, puisqu'il est possible de construire de nouveaux noms composés à partir des noms composés déjà existants.

Des 4 200 noms composés recensés grâce à cette méthode, 800 correspondent à des termes de spécialité (cf. III, 6.7), qui se répartissent en soixante-deux domaines de spécialité, à savoir : ecclésiastique, beaux-arts, militaire, philosophie, physique, aéronautique, sport, anatomie, archéologie, astrologie, astronomie, biologie, biochimie, balistique, botanique, écologie, géographie, géologie, géométrie, linguistique, grammaire, droit international, théologie, sciences, électricité, électronique, religion, médecine, histoire, cinéma, sociologie, folklore, logique, comptabilité, mathématiques, météorologie, mécanique, ingénierie, musique, maritime, droit, droit civil, immobilier, économie, optique, anthropologie, informatique, rhétorique, physiologie, chimie, psychologie, politique, statistique, technique, technologie, télécommunications, topographie, pharmacologie, photographie, psychanalyse, psychiatrie, océanographie. A titre indicatif, citons les exemples suivants :

μανία καταδίωξης

(manie de persécution) (domaine : Psychologie et Médecine)

λεκάνη απορροής

(bassin de réception) (domaine : Géologie)

ταχύτητα του ήχου

(vitesse de son) (domaine : Physique)

μάρτυρας κατηγορίας

(témoin à charge) (domaine : Droit)

πυραμίδα του πληθυσμού

(pyramide de la population) (domaine : Sociologie)

συμπλήρωμα γωνίας

(supplément d'angle) (domaine : Mathématiques)

αγορά εργασίας

(marché du travail) (domaine : Economie)

δύναμη κρούσης

(force de frappe) (domaine : Militaire)

δρόμος αντοχής

(course de fond) (domaine : Sport)

Parmi les noms composés recensés manuellement à partir des dictionnaires usuels, 739 correspondent à des entités nommées (cf. III, 6.6). On désigne traditionnellement par le terme d'« entité nommée » les noms propres, mais aussi les expressions de temps et de quantité (cf. C. Martineau *et al.* 2007). Le terme d'« entité nommée » est utilisée dans le cadre du TAL depuis les conférences *Message Understanding Conferences* (MUC⁶⁰). Plusieurs typologies ont été proposées pour les entités nommées (cf. C. Le Meur ; S. Galliano et E. Geoffrois 2004, S. Sekine *et al.* 2002). Dans le cadre de cette étude nous adoptons la typologie proposée par (C. Martineau *et al.* 2007) (cf. aussi III, 6.6). Elle comporte notamment les classes d'entités nommées suivantes :

Noms de personnes

Noms d'organisations

Noms de lieux

Noms de faits

Noms d'œuvres

Noms de moyens

Citons ici quelques exemples d'entités nommées recensées à partir des dictionnaires et des grammaires d'usage :

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

(Université Aristote de Thessalonique) (entité nommée d'« organisation »)

Αφροδίτη της Μήλου

(Vénus de Milo) (entité nommée d'« œuvre »)

γέφυρα των Στεναγμών

(pont des Soupirs) (entité nommée de « lieu »)

Γη του Πυρός

(Terre de Feu) (entité nommée de « lieu »)

Δεσποτάτο της Ηπείρου

(Despotat d'Épire) (entité nommée de « lieu »)

Δικηγορικός Σύλλογος Αθήνας

(Ordre des Avocats d'Athènes) (entité nommée d'« organisation »)

Ευρωπαϊκή Ζώνη Ελευθέρων Συναλλαγών

(Association européenne de libre-échange) (entité nommée d'« organisation »)

Ευρωπαϊκή Τράπεζα Επενδύσεων

(Banque européenne d'investissement) (entité nommée d'« organisation »)

Πλατεία Ομονοίας

(Place de la Concorde) (entité nommée de « lieu »)

Ψυχιατρικό Νοσοκομείο Αθήνας

(Hôpital Psychiatrique d'Athènes) (entité nommée d'« organisation »)

La méthode manuelle de recensement nous a permis de rendre compte et d'examiner en profondeur certains phénomènes spécifiques des noms composés de structure $N(E + DET:G) N:G$, puisqu'ils étaient déjà décrits dans les dictionnaires et les grammaires d'usage. Nous abordons tous ces phénomènes dans ce qui suit.

Nous avons identifié des séquences rares où le composant nominal au génitif (simple ou composé) – qui est canoniquement postposé – est obligatoirement antéposé par rapport au nom-tête :

$N:G N$

φόρου υποτελής

taxe-_{Gms} tributaire

**υποτελής φόρου*

**tributaire taxe-_{Gms}*

(tributaire)

⁶⁰ Voir le site : <http://www.muc.saic.org>

N:G N

διαβόλου (σπορά + γέννα + κάλτσα)

diabole-G_{ms} (germe + accouchement + chaussette)

**(σπορά + γέννα + κάλτσα) διαβόλου*

**(germe + accouchement + chaussette) diabole-G_{ms}*

(un bon petit diable)

N:G N

χαράς Ευαγγέλια

joie-G_{fs} Evangiles

**Ευαγγέλια χαράς*

**Evangiles joie-G_{fs}*

(bonnes nouvelles)

DET:G N:G N

ενός λεπτού σιγή

une-G_{ns} minute-G_{ns} silence

**σιγή ενός λεπτού*

**silence un-G_{ns} minute-G_{ns}*

(une minute de silence)

Les cas de figure, que nous venons de présenter, constituent des noms composés obligatoirement permutés. Notons que, lorsque le nom au génitif est déterminé par l'article défini, le nom-tête reçoit lui-aussi la détermination définitive. Il s'agit du phénomène de redoublement de l'article défini, propre à la langue grecque (I. Tsamadou 1984). Par exemple :

*DET+Déf:G N:G (*E + DET+Déf) N*

*του λιναριού (*E + τα) (πάθη + βάσανα)*

le-G_{ns} lin-G_{ns} (*E + les) (souffrances + épreuves)

(grandes souffrances)

*DET+Déf:G N:G (*E + DET+Déf) N*

*του κόλου (*E + τα) εννιάμερα*

le-G_{ns} cul-G_{ns} (*E + les) neufs jours

(sottises)

*DET+Déf:G N:G (*E + DET+Déf) N*

*της Άρτας (*E + το) γιοφύρι* (deux entrées : nom composé et entité nommée)

*la-Gfs Arta-Gfs (*E + le) pont*

(le pont d'Arta + œuvre inachevée)

*DET+Déf:G N:G (*E + DET+Déf) N*

*του πουλιού (*E + το) γάλα*

*le-Gns oiseau-Gns (*E + le) lait*

(abondance de mets)

Notons que les noms composés obligatoirement permutés constituent souvent des syntagmes nominaux figés faisant partie d'une phrase simple figée. Cependant, le fait qu'ils ont un emploi autonome, en dehors de l'expression figée, justifie notre choix de les inclure dans notre inventaire.

D'autres cas de figure mettent en évidence le problème de limite entre noms composés et syntagmes nominaux semi-figés (ou productifs). Il s'agit des constructions dont le composant nominal au génitif s'applique à un large spectre des noms. A titre indicatif, citons le paradigme du *N:G* : *πολυτελείας*/de luxe :

είδος πολυτελείας

(article de luxe)

εστιατόριο πολυτελείας

(restaurant de luxe)

κατάστημα πολυτελείας

(boutique de luxe)

ξενοδοχείο πολυτελείας

(hôtel de luxe)

τηλεγράφημα πολυτελείας

(télégramme de luxe)

φόρος πολυτελείας

(taxe de luxe)

χαρτί πολυτελείας
(papier de luxe)

ψωμί πολυτελείας
(pain de luxe)

Dans la majorité des cas, du point de vue syntaxique, le composant nominal *πολυτελείας*/luxe est complément du nom-tête :

Η πολυτέλεια του ξενοδοχείου εντυπωσιάζει το Νίκο
(Le luxe de l'hôtel impressionne Nicolas)

Dans certains cas, il semble exercer une fonction adjectivale, ce qui justifie la mise en relation avec la structure AN^{61} :

$N\ N:G = N:G - a\ N$
ξενοδοχείο πολυτελείας = πολυτελές ξενοδοχείο
(hôtel de luxe)

Dans d'autres cas, la mise en relation avec la structure AN n'est pas possible :

**πολυτελές τηλεγράφημα*
luxueux télégramme
(télégramme de luxe)

Du point de vue sémantique, le composant nominal au génitif permet d'assigner une qualification au nom-tête, en l'occurrence, la très haute qualité d'un produit ou d'un service, sur la base de laquelle se forme également son prix. Notons, enfin, que parfois la permutation des composants est possible : $N:G\ N = :$ *πολυτελείας ξενοδοχείο*/hôtel de luxe.

Citons ici d'autres noms au génitif (avec ou sans déterminant) qui forment des combinaisons nominales comme celles que nous venons de présenter :

- (*χώρα + παιδεία + Ευρώπη + ...*) (*Ε + των*) *δύο ταχυτήτων*
(pays + éducation + Europe + ...) (*E + les-G_{fp}*) *deux-G_{fp} vitesses-G_{fp}*
((pays + éducation + Europe + ...) (à deux vitesses))

⁶¹ Pour la mise en relation des $N\ (E+DET:G)\ N:G$ avec les AN cf. II, 5.2.5.5 ; V, 11.2.1 et 11.3.3.

pour assigner au nom-tête la qualification de l'évolution inégale ;

- *(προϊόν + ταινία + όχημα + ...)* διαλογής
(produit + film + véhicule + ...) choix-Gfs
((produit + film + véhicule + ...) (de second choix))
pour des produits de second choix, avec des défauts, qui sont vendus à un prix très bas ;
- *(ρολόι + ρούχα + καφετιέρα + ...)* μάρκας
(montre + vêtements + cafetière + ...) marque-Gfs
((montre + vêtements + cafetière + ...) (de marque))
pour les produits de marques connues ;
- *(ευκαιρία + βιντεοπαιχνίδια + υπηρεσίες + ...)* πρώτης τάξεως
(occasion + jeux vidéo + services + ...) premier-Gfs ordre-Gfs
((occasion + jeux vidéo + services + ...) (de premier ordre))
pour quelque'un ou quelque chose d'exceptionnel ;
- *(δίαιτα + παντελόνι + γλώσσα + ...)* της μόδας
(régime + pantalon + langue + ...) la-Gfs mode-Gfs
((régime + pantalon + langue + ...) (à la mode))
pour quelque chose qui est à la mode.

Toutefois, le phénomène inverse s'observe aussi. Il arrive qu'un certain nombre des noms puisse se combiner avec un large spectre des noms au génitif. En voici quelques exemples :

- *(ζήτημα + θέμα)* (τιμής + χρημάτων + αρχών + ...)
(question (d'honneur + d'argent + de principes + ...))
pour désigner que quelque chose dépend surtout ou exclusivement de ce qui est désigné par le nom au génitif ;
- *τέρας* (μνήμης + ευφυΐας + υπομονής + αντοχής + ασχήμιας + ...)
(monstre (de mémoire + d'intelligence + de patience + de tolérance + de laideur + ...))
pour désigner une personne qui a une qualité importante qui la caractérise, désignée par le nom au génitif ;

- *άνεμος ((E + της) αλλαγής + (E + του) πολέμου + (E + της) ανανέωσης + (E + της) ελευθερίας +...)*
(vent ((de + du) changement + (de + de la) guerre + (de + du) renouvellement + (de + de la) liberté +...))
pour désigner une tendance forte qui présage ce qui est désigné par le nom au génitif ;
- *άνθρωπος (του θεάματος + του θεάτρου + των γραμμάτων και των τεχνών + των άκρων + ...)*
(personne (du spectacle + du théâtre + des lettres et des arts + des extrêmes +...))
pour désigner la qualité d'une personne ou le domaine professionnel auquel elle appartient, ou bien
- *άνθρωπος (του υπουργού + του σπιτιού + του καταστήματος +...)*
(personne (du ministre + de la maison + du magasin + ...))
pour désigner la personne qui est au service de ce qui est désigné par le nom au génitif ou la personne confiante ou intime de ce qui est désigné par le nom au génitif.

Dans les deux cas, les listes de ces noms peuvent être très vastes et doivent rester ouvertes. Quant à leur représentation formalisée et reconnaissance lexicale automatique, nous avons donc choisi de les représenter à l'aide de graphes lexicalisés construits manuellement (cf. V, 11.3.2). C'est d'ailleurs la solution que nous avons adoptée pour la représentation formalisée et la reconnaissance lexicale automatique des syntagmes nominaux semi-figés (ou productifs) dont le spectre du nom-tête ou du deuxième composant nominal est supérieur à dix.

Toutefois, la compilation manuelle des dictionnaires et des grammaires présente certaines limites. Tout d'abord, il s'agit d'une méthode coûteuse. Le temps qui s'est avéré nécessaire pour que nous effectuions la compilation manuelle des dictionnaires et des grammaires d'usage correspond à 9 hommes-mois. Soulignons que, grâce à cette méthode, nous n'avons dressé qu'une première liste qui est loin d'être exhaustive, car nous nous sommes limitée à la compilation de dictionnaires de la langue générale. La compilation de dictionnaires spécialisés serait indispensable afin de compléter cette liste. Il devient donc évident que cette méthode de recensement manuelle constitue un projet de longue haleine.

Le nombre d'heures important consacré à la compilation manuelle est essentiellement dû à des problèmes qui touchent à la description des noms composés dans les dictionnaires usuels. Dans la plupart des cas, les noms composés ne constituent pas d'entrées autonomes dans les

dictionnaires usuels. Le dictionnaire d'A. Anastassiadis-Syméonidis (2002) en est la seule exception, dans lequel un statut autonome est accordé aux noms composés. Les noms composés sont, en général, représentés de manière très diverse et il n'est donc pas toujours facile de les repérer. Ils figurent soit sous l'entrée du nom-tête soit sous l'entrée du composant nominal au génitif et leur codage est loin d'être clair et systématique. Ils comportent parfois une étiquette attestant de leur figement : ils sont qualifiés tantôt d'expressions (notés *εκφρ.*) tantôt de locutions (notés *ΦΡ*). Mais, dans la majorité des cas, ils ne comportent aucune étiquette et, de manière générale, ils ne sont jamais qualifiés de noms composés. Nous les retrouvons souvent dans le corps d'article ou dans celui des exemples portant sur l'entrée principale.

Il arrive quelquefois qu'ils soient enregistrés uniquement au pluriel (e.g. *εγκλήματα πολέμου*/crimes de guerre), mais aucune explication n'est donnée pour justifier ce choix ; ce qui prête à confusion quant à la prise en compte des contraintes morphologiques lors du traitement automatique des noms composés en question. De plus, ils sont représentés tantôt avec l'article défini (e.g. *Τάγμα της Τιμής*/Bataillon de l'Honneur) tantôt sans article défini (e.g. *Τάγμα Τιμής*/Bataillon d'Honneur).

De manière générale, la flexion des noms composés n'est pas décrite et les informations syntaxiques et sémantiques (notamment les définitions) les concernant sont absentes des dictionnaires usuels. De plus, les noms composés ne sont pas systématiquement illustrés par des exemples d'emploi dans une phrase simple, ce qui empêche de rendre compte de leur comportement syntaxique et des relations qu'ils entretiennent avec les autres constituants de la phrase. Ce problème devient plus évident quand il s'agit d'identifier les noms composés prédicatifs ou encore les syntagmes nominaux figés faisant partie des phrases simples figées. Par exemple, seule la séquence nominale *σημείο ζωής*/signe de vie figure dans les dictionnaires, alors que sa combinatoire unique avec les verbes supports *δίνω*/donner et *παίρνω*/recevoir n'y est pas représentée⁶².

Du point de vue de la constitution lexicale interne, les informations concernant les possibilités de variation des noms composés ne figurent pratiquement jamais dans les ouvrages usuels. Dans le cadre fortement descriptif et classificatoire où se situe cette étude, il est important de connaître la structure lexicale maximale dans laquelle entre chaque nom composé ainsi que

⁶² Sur les limites entre expressions à éléments figés et expressions à verbe support cf. III, 6.8. Sur l'étude des noms prédicatifs cf. IV, 9.1-9.3.

toutes ses possibilités de variation. Repérer donc les variantes soudées ou encore les variantes de structure (par exemple, les formes réduites) reste une tâche ardue qui n'est facilitée en rien par la consultation des dictionnaires et des grammaires édités (cf. II, 5.2.5.1-5.2.5.8 et II, 5.3.4).

Quant aux néologismes (e.g. *γρίπη των πτηνών*/grippe aviaire), ceux-ci ne sont pas listés dans les dictionnaires d'usage, à l'exception du dictionnaire d'A. Anastassiadis-Syméonidis (2002). Du point de vue du TAL, la reconnaissance de tous les mots ainsi que de toutes les combinaisons possibles de mots semble être la préoccupation primordiale (cf. L. Danlos 1986).

Comme le note également A. Poncet-Montange (1991 : 34), un autre facteur d'incomplétude des dictionnaires usuels tient au fait qu'une grande partie des noms composés sont des termes techniques. Or, la capacité d'accueil des termes dans un dictionnaire d'usage de langue générale est limitée. Il faudra donc procéder au dépouillement de dictionnaires encyclopédiques et de textes variés afin d'augmenter le degré de couverture « lexicale » d'un dictionnaire électronique comme le DELAC.

Dans le but d'augmenter la couverture « lexicale » de notre dictionnaire, nous avons également consulté des sources spécialisées, telles que les bases terminologiques d'ELETO (Société Hellénique de Terminologie). ELETO met à la disposition du public des bases terminologiques concernant les trois domaines suivants : télécommunications⁶³, technologie de l'information⁶⁴ et terminologie⁶⁵. A titre indicatif, citons quelques exemples de termes tirés des bases terminologiques d'ELETO :

γραμμή αναβατών

(ligne des ascendantes) (domaine : Télécommunications)

γραμμή καταβατών

(ligne des descendantes) (domaine : Télécommunications)

σύστημα μετάβασης

(système de transition) (domaine : Technologie de l'information)

⁶³ http://www.moto-teleterm.gr/search_gr.asp

⁶⁴ <http://inforterm.cs.aueb.gr/greek/search.php>

⁶⁵ <http://www.eleto.gr/gr/termbases.htm>. Précisons qu'il s'agit d'une base terminologique consacrée à la discipline de la terminologie.

εξελεγκτήρας τεχνοτροπίας

(correcteur stylistique) (domaine : Technologie de l'information)

τεχνική των ημιαγωγών

(technique des semiconducteurs) (domaine : Technologie de l'information)

φύλλο τεχνοτροπίας

(feuille de style) (domaine : Technologie de l'information)

ακολουθία χαρακτήρων

(séquence de caractères) (domaine : Terminologie)

εξέλεγχος συνέπειας

(contrôle de cohérence) (domaine : Terminologie)

παράγοντας στάθμισης

(facteur de pondération) (domaine : Terminologie)

Il est clair que la création continue de noms composés techniques défie le traitement automatique des textes et complique l'étude linguistique des noms composés. Les termes de spécialité présentent des problèmes qui leur sont spécifiques et que l'on est pas en mesure de résoudre en tant que non-spécialiste de chaque domaine. Par exemple, on n'est pas en mesure de connaître si un terme de spécialité présente des variantes abrégées et de les répertorier. De plus, on ne peut pas être certains des éventuelles restrictions de flexion (e.g. est-ce que le terme *γραμμή αναβατών*/ligne des ascendantes a un pluriel aussi ?) ou des variantes de structure (e.g. pour *γραμμή αναβατών*/ligne des ascendantes, est-ce qu'il y a une variante avec déterminant : *γραμμή των αναβατών*/ligne des ascendantes ?).

Contrairement aux autres unités linguistiques, les termes tendent à présenter un rapport univoque entre une notion (définie au préalable) et sa dénomination dans un domaine donné. Le classement, l'étude des procédés de formation, la normalisation et la diffusion des termes font d'ailleurs l'objet d'une discipline spécifique. On peut donc se demander si une étude générale sur la composition nominale peut être menée sans la participation de terminologues.

A propos des termes de spécialité, G. Gross (1996a : 27) remarque que « la terminologie a été un des vecteurs de l'étude des noms composés. Si l'on fait l'inventaire du vocabulaire des

langues de spécialité, on se rend compte que les substantifs constituent l'essentiel du vocabulaire et que parmi eux les noms composés s'y taillent la part du lion ».

Notons ici que l'objet d'un dictionnaire électronique morphologique général décrit la morphologie et la flexion des noms composés de la langue courante. En ce qui concerne son enrichissement avec des termes de spécialité, il y aurait un choix à faire : les termes de spécialité peuvent être inclus dans ce dictionnaire ou bien ils pourraient constituer l'objet des dictionnaires électroniques spécialisés que les utilisateurs appliqueraient en fonction du contenu plus ou moins technique d'un texte. Pour ce qui est de notre travail, nous avons pris en compte un nombre très limité de termes de spécialité, que nous avons repérés dans les sources précitées. Nous n'avons pas effectué des recherches supplémentaires sur l'éventuel statut terminologique des autres séquences recensées. Nous n'avons qualifié de 'termes de spécialité' que les séquences qui étaient désignées en tant que tels par les sources documentaires elles-mêmes. Nous leur avons assigné le code TS dans les tables du lexique-grammaire et nous proposons également l'ajout de ce code dans le dictionnaire électronique morphologique pour les entrées qui constituent des termes de spécialité.

De cette façon, nous avons obtenu une première liste de séquences $N(E + DET:G)N:G$ qui comportent des degrés de figement divers. Malgré les lacunes et les omissions, cette première liste nous a permis d'entamer l'étude de cette partie du lexique et d'en déduire un certain nombre de régularités et de spécificités. Enfin, cette méthode nous a permis entre autres de recenser des noms composés qui n'apparaissent que rarement dans les corpus et qui ne peuvent donc pas être repérés par des méthodes statistiques.

4.2.3 Méthode semi-automatique : application de motifs morphosyntaxiques à des corpus de textes écrits

Bien qu'elles constituent un excellent point de départ pour le recensement des formes composées, les sources lexicographiques sont incomplètes. Le défaut des dictionnaires et des grammaires édités peut être compensé par d'autres pistes, telles que le dépouillement de corpus de textes.

Le corpus de textes numérique, sur lequel nous nous sommes fondée pour la constitution de notre liste de noms composés $N(E + DET:G)N:G$, est un corpus de presse (journal intitulé *TA NEA*, année 2003), de taille 21,4 Mo, ce qui représente environ 1 500 000 occurrences de mots. Nous nous sommes contentée de ce corpus de taille relativement petite car il nous a

fourni un nombre de candidats noms composés suffisant pour la présente étude. Il s'agit d'un corpus qui a été utilisé dans le cadre d'un projet de recherche national⁶⁶ concernant les technologies du langage. Le corpus a été prétraité à l'aide du système d'analyse automatique des textes *Unitex*⁶⁷ (S. Paumier 2003). Le prétraitement effectué comprend les opérations suivantes, présentées ici selon l'ordre d'application : le découpage du texte en phrases – qui a été évalué manuellement dans son ensemble, le découpage des phrases en unités lexicales et l'analyse lexicale du texte à l'aide des dictionnaires des mots simples (DELAF) et composés (DELACF). Notons que le corpus a été annoté manuellement avec des informations issues de la linguistique de corpus (cf. K. Ioannidou *et al.* 2005). En gros, il s'agit des informations sur l'auteur, la date de parution, la source d'édition, le genre textuel ainsi que sur la macrostructure et la microstructure des textes. A titre d'illustration, citons ci-dessous un petit extrait du corpus utilisé :

```

BeginArticle
<POLITIS>Ανεπτυγμένη είδηση-αφήγηση</POLITIS>
<CATEGORY>ΕΛΛΑΔΑ</CATEGORY>
<H2>{S}Κοκτέιλ θανάτου για 3 κρατούμενες</H2>
<H4>{S}Κύκλωμα διακινούσε ανενόχλητα ναρκωτικά και ψυχοτρόπα στις Γυναικείες
Φυλακές
Κορυδαλλού </H4>
<AUTHOR>{S}ΛΙΑ ΝΕΣΦΥΓΕ
{S}ΠΡΟΚΟΠΗΣ ΓΙΟΓΙΑΚΑΣ</AUTHOR>

<LEAD>{S}Ανενόχλητα και κάτω από τη μύτη των σωφρονιστικών υπαλλήλων, κύκλωμα
από κρατούμενες διακινούσε ηρεμιστικά και αντικαταθλιπτικά χάπια, έναντι
αδρών ανταλλαγμάτων, στις Γυναικείες Φυλακές του Κορυδαλλού.</LEAD>

{S}Ο θάνατος τριών και η μεταφορά άλλης μίας κρατούμενης, σε κόμα, στο
νοσοκομείο, ύστερα από χρήση θανατηφόρου κοκτέιλ με ηρωίνη και χάπια, έφερε
με τραγικό τρόπο ξανά στο προσκήνιο αυτό που είναι κοινό μυστικό, ότι δηλαδή
στις φυλακές γίνεται διακίνηση ναρκωτικών και ψυχοτρόπων...

{S}Επικεφαλής του κυκλώματος, σύμφωνα με την Ασφάλεια, είναι η Μαριάνθη
Πατσέλη, γνωστή για την υπόθεση ναρκωτικών με τον τραγουδιστή των Νοστράδαμος
Ιπποκράτη Εξαρχόπουλο (είχαν συλληφθεί το 1996 με 10 κιλά ηρωίνη).{S} Στο
κύκλωμα φέρεται ότι συμμετείχε και άλλη μία κρατούμενη, η οποία
αποφυλακίστηκε πρόσφατα.

{S}Η ίδια με συνεργό τη συγκρατούμενη της στο ίδιο κελί, τη Ρωσίδα Όλγα
Βελιγκανόβα, έτριβαν τα χάπια και τα μετέτρεπαν σε δόσεις (φιξάκια), τα οποία
μοίραζαν με το αζημίωτο από κελί σε κελί!{S} Όσες ήθελαν να τα προμηθευτούν
έδιναν χρυσάφια, τηλεκάρτες, γυαλιά ηλίου και φυσικά χρήματα, όπως
αποδεικνύεται από τα ευρήματα που κατασχέθηκαν στο κελί τους...

{S}Σύμφωνα με την έως τώρα έρευνα, τα τέσσερα θύματα είχαν προμηθευτεί από
τρία φιξάκια με χάπια η καθεμία, ενώ το θανατηφόρο κοκτέιλ κατασκευάστηκε με
την πρόσμειξη τριών γραμμαρίων ηρωίνης.

<H6>{S}Η μητέρα της</H6>

```

⁶⁶ Projet de recherche national coordonné par le Centre pour la Langue Grecque (KEΓ), qui porte sur l'amélioration, l'extension et la création d'un corpus numérique en grec moderne.

⁶⁷ www-igm.univ-mlv.fr/~unitex/

{S}Την ηρωίνη, όπως διαπιστώθηκε, την πέρασε στη φυλακή, στις 30 Ιανουαρίου, στο απογευματινό επισκεπτήριο η μητέρα της κρατούμενης Μαρίας-Σοφίας Μπακέα.{S} Η 59χρονη Ντόρα Κάρμεν Βιγιάμπα, πρόκειται να οδηγηθεί σήμερα στον εισαγγελέα Πειραιά που διενεργεί την προανάκριση κ. Γρηγόρη Πεπόνη, μαζί με τον φίλο της κόρης της και πατέρα του ανήλικου παιδιού της, τον Δημήτρη Πενιά.{S} Την ηρωίνη την είχε η 59χρονη από την Αργεντινή στη χούφτα της και την ώρα που χαιρετούσε την κόρη της την έβαλε με τρόπο στην μπλούζα της χωρίς να γίνει αντιληπτή.

{S}Το κουβάρι άρχισε να ξετυλίγεται όταν η 23χρονη Στεφανία Ρουμπέτη, η οποία βρέθηκε σε κώμα και νοσηλεύεται εκτός κινδύνου στο Γενικό Κρατικό Νοσοκομείο της Νίκαιας, αποκάλυψε, σύμφωνα με πληροφορίες, ότι την ηρωίνη την έφερε στις φυλακές η 42χρονη Μαρία-Σοφία Μπακέα, η οποία είναι η μοναδική από τις πέντε κρατούμενες που δεν έπαθε το παραμικρό.{S} Και αυτό διότι, σύμφωνα με πληροφορίες από τις φυλακές, δεν έκανε χρήση του κοκτέιλ, αλλά πήρε μόνο ηρωίνη.

<REF>
 <SOURCE>TA NEA</SOURCE>
 <DATE>02-01-2003</DATE>
 <CAT>Σελ.: N10</CAT>
 <ED>A17529N101</ED>
 <ID>ID:347746</ID>
 <URL>http://ta-
 nea.dolnet.gr/neaweb/neafire.print_unique?e=A&f=17529&m=N10&aa=1</URL>
 </REF>⁶⁸

Nous nous sommes servie des fonctionnalités du système *Unitex* (S. Paumier 2003) afin de compléter – de manière semi-automatique – la liste des noms composés issue de la compilation des grammaires et des dictionnaires. Plus précisément, nous avons effectué des requêtes de motifs morphosyntaxiques exprimés sous forme d'expressions régulières (ou rationnelles). Les expressions régulières (ou rationnelles) permettent de repérer des unités lexicales au moyen d'un jeu de catégories grammaticales et de filtres morphologiques qui

⁶⁸ Ci-dessous, nous explicitons le contenu des balises figurant dans l'exemple du corpus donné.

<POLITIS> : genre textuel, qui fut ajouté après avoir appliqué des critères de classification issus de la linguistique de corpus.

<CATEGORY> : Nom de la rubrique du journal dans laquelle apparaît l'article.

<H2> : Présence de « titre de base ». Notons que tous les textes (ou articles) sont accompagnés d'un « titre de base ».

<H3> : Présence de « titre intermédiaire ». Notons que les textes (ou articles) ne sont pas toujours accompagnés de « titres intermédiaires ».

<H4> : Présence de « sous-titre ». Notons que les textes (ou articles) ne sont pas toujours accompagnés de « sous-titres ».

<AUTHOR> : Nom ou noms de l'auteur ou des auteurs du texte (ou article). Notons que les textes (ou articles) ne sont pas toujours accompagnés du nom de l'auteur. Dans ce cas, nous précisons que l'article est « non-signé ».

<LEAD> : Présence d'un paragraphe spécial ou, dans quelques cas, d'une seule phrase. Il joue le rôle de résumé du texte (ou article) ou d'introduction, dont le but est d'attirer l'intérêt du lecteur.

<H6> : Présence de « titre intermédiaire », dont le contenu d'informations est moins important que celui du « titre intermédiaire » H3.

<REF> : Description des « informations peritextuelles » qui accompagnent un texte (ou article). Ces informations comprennent la source (ou nom du journal), la date de parution, le code d'identification (ID) et le site web du texte (ou article). Notons, enfin, que tous les textes (ou articles) sont obligatoirement accompagnés de leurs « informations peritextuelles ».

tiennent compte des contraintes flexionnelles (S. Paumier 2006). A titre illustratif, citons quelques exemples des séquences $N (E + DET:G) N:G$ tirées du corpus numérique :

Οργανισμός Ασφάλισης Ελεύθερων Επαγγελματιών

(Organisation d'Assurance de Travailleurs Indépendants) (entité nommée d'« organisation »)

κάρτα τεχνικού ελέγχου

(carte de contrôle technique)

Άρμα Θεάτρου

(Arma Theatrou) (entité nommée d'« organisation »)

ψηφιακή τεχνολογία απεικόνισης

(technologie numérique de représentation)

Précisons que l'application d'une formule simple s'est avérée suffisante pour obtenir des noms composés complexes (ou surcomposés) aussi, car il existe des noms composés qui sont déjà reconnus grâce à l'application du DELACF existant.

Notre intuition initiale était que les noms $NN:G$ et les noms $NDET:G N:G$ ont des comportements syntaxiques différents et constituent, éventuellement, deux types de noms composés distincts. Pour cette raison, nous avons choisi d'appliquer deux motifs morphosyntaxiques, à savoir $\langle N \rangle \langle N:G \rangle$ et $\langle N \rangle DET \langle N:G \rangle$. L'application de tels motifs morphosyntaxiques est possible grâce au dictionnaire des formes fléchies simples qui est assez complet. De cette manière, nous avons pu obtenir deux listes de candidats noms composés, que nous avons par la suite confrontées. La liste des candidats $NN:G$ comporte 50 564 occurrences, alors que celle des candidats $NDET:G N:G$ comporte 59 198 occurrences.

Notre intuition n'a été vérifiée qu'en partie, car nous avons constaté que, dans environ 10% des cas, les séquences sans déterminant étaient des variantes réduites des séquences avec déterminant. L'étude des variantes de déterminant est présentée en détail dans II, 5.2.5.1. A titre d'illustration, citons les exemples suivants :

Μνημείο (Ε + του) Άγνωστου Στρατιώτη

monument (E + le-_{Gms}) inconnu-_{Gms} soldat-_{Gms}

(tombe du soldat inconnu)

συνήγορος (Ε + της) πολιτικής αγωγής

avocat (E + la-*Gfs*) civile-*Gfs* action-*Gfs*

(avocat de la partie civile)

δώρο (Ε + των) Χριστουγέννων

cadeau (E + les-*Gnp*) Noël-*Gnp*

(treizième salaire)

χάσμα (Ε + των) γενεών

fossé (E + les-*Gfp*) générations-*Gfp*

(conflit de générations)

A partir de ces deux listes et, après application de nos critères de délimitation des noms composés $N(E + DET:G)N:G$ (cf. I, 2.2.1-2.2.11), nous n'en avons retenu que 6 000, dont une grande partie avait déjà été recensée grâce à la méthode manuelle (cf. II, 4.2.1). Après avoir éliminé les doublons, nous n'en avons retenu que 5 000 nouveaux noms composés, dont environ 1 400 entités nommées et environ 3 400 syntagmes nominaux semi-figés (ou productifs). Il faut noter que la méthode semi-automatique est aussi coûteuse que la méthode manuelle, car elle comprend une étape considérable de validation manuelle, qui correspondrait à 9 hommes-mois.

La liste brute des candidats noms composés $N(E + DET:G)N:G$ contenait un nombre important d'erreurs ainsi que de séquences libres (ou compositionnelles). Dans un premier temps, nous avons éliminé les séquences libres à l'aide des critères de délimitation des noms composés (cf. I, 2.2.1-2.2.11) et en nous appuyant sur notre compétence en tant que locuteur natif du grec moderne. Citons ici quelques exemples de séquences libres (ou compositionnelles) présentes dans la liste brute :

ανυπαρξία βιβλιοθηκών

(manque de bibliothèques)

επιλογή γραβάτας

(choix de cravate)

σταγόνες πετρελαίου

(gouttes de pétrole)

γιος του προέδρου

(fils du président)

δημιουργία ενός κλαμπ

(création d'un club)

νεφροί του μωρού

(reins du bébé)

τσάντα του θύματος

(sac de la victime)

Dans un deuxième temps, nous avons identifié des formes, dont seulement une partie figurait dans la liste brute, que nous avons par la suite complétées. Sur une base initialement intuitive, nous avons observé que souvent la forme contenue dans la liste faisait éventuellement partie d'une séquence plus longue, correspondant soit à une entité nommée soit à un nom composé commun. Nous avons effectué des recherches supplémentaires sur les tables des concordances obtenues par *Unitex* et sur le Web afin de vérifier notre hypothèse et compléter la partie manquante de la séquence obtenue. Notre hypothèse a été vérifiée tant pour les entités nommées que pour les noms composés communs. A titre d'illustration, citons les exemples suivants :

Σεquences incomplètes obtenues	Σεquences complètes
<i>Ένωση Αερομεταφορών</i> (Association du transport aérien)	<i>Διεθνής Ένωση Αερομεταφορών</i> (Association internationale du transport aérien) (entité nommée d'« organisation »)
<i>Ένωση Αγαθοεργίας</i> (Union de Charité)	<i>Γενική Αρμενική Ένωση Αγαθοεργίας</i> (Union Arménienne Générale de Charité) (entité nommée d'« organisation »)
<i>Ένωση Δήμων</i> (Union de Municipalités)	<i>Κεντρική Ένωση Δήμων και Κοινοτήτων Ελλάδας</i> (Union Centrale des Municipalités et des Communautés de Grèce) (entité nommée d'« organisation ») <i>Τοπική Ένωση Δήμων και Κοινοτήτων</i>

	(Union Locale de Municipalités et de Communautés) (entité nommée d'« organisation »)
<i>Ιδρυμα Ραδιοφωνίας</i> (Fondation de Radiophonie)	<i>Εθνικό Ιδρυμα Ραδιοφωνίας</i> (Fondation Nationale de Radiophonie) (entité nommée d'« organisation »)
<i>ανιχνευτής χημικών</i> (détecteur chimique)	<i>ανιχνευτής χημικών ουσιών</i> (détecteur chimique)
<i>αμοιβαία κεφάλαια διαχείρισης</i> (fonds commun de gestion)	<i>αμοιβαία κεφάλαια διαχείρισης διαθεσίμων</i> (fonds commun de gestion de réserves)

En ce qui concerne les erreurs présentes dans la liste brute, celles-ci sont dues essentiellement aux formes ambiguës (cf. aussi V, 11.3.1). A titre d'illustration, citons les exemples suivants, qui n'ont pas été retenus dans la liste finale :

$(A_{-Gnp} + N_{-Gnp}) N_{-Gnp} = :$ αστικών κέντρων
(agglomération urbaine)

La forme *αστικών* correspond à la fois au nom (adjectif substantivé) *αστικό*/bus (codé dans le DELAF en tant que *N*) et à l'adjectif *αστικό*/urbain (codé dans le DELAF en tant que *A*).

$N_{-Nfs} (N_{-Nns} + V_{-p3s}) = :$ καινοτομία είναι
(innovation est)

La forme *είναι* correspond à la fois au nom *είναι*/être (codé dans le DELAF en tant que *N*) et à la troisième personne du présent de l'indicatif du présent du verbe *είμαι*/être (codé dans le DELAF en tant que *V*).

$N_{-Nnp} (N_{-Gfs} + A_{-Gfs}) = :$ κείμενα μικρής
(textes de petite)

La forme *μικρής* correspond à la fois au nom (adjectif substantivé) *μικρή*/jeune (codé dans le DELAF en tant que *N*) et à l'adjectif *μικρή*/jeune (codé dans le DELAF en tant que *A*).

$N_{-Gfs} (N_{-Gns} + PRO+POSS_{-Gfs}) = :$ σύνδεσή σου
(ta connexion)

La forme *σου* correspond à la fois au nom *σου*/chou à la crème (codé dans le DELAF en tant que *N*) et au pronom possessif *σου*/ta (codé dans le DELAF en tant que *PRO*).

Cependant, dans un certain nombre des cas, ces formes constituent des noms composés et sont donc retenues dans la liste finale, comme par exemple :

κινητό νέας γενιάς

(portable de nouvelle génération)

κινητό τρίτης γενιάς

(portable de la troisième génération)

La forme *κινητό* correspond à la fois à l'adjectif *κινητό*/portable (codé dans le DELAF en tant que *A*) et au nom (adjectif substantivé) *κινητό*/téléphone portable (codé dans le DELAF en tant que *N*). Apparemment, *Unitex* a généré ces séquences car il a reconnu *νέας γενιάς*/nouvelle génération et *τρίτης γενιάς*/troisième génération en tant que noms composés grâce à l'application du DELACF actuel. Dans un souci d'exhaustivité et afin de ne pas perdre d'informations pertinentes pour notre tâche lexicographique, nous avons validé manuellement l'ensemble des formes ambiguës.

Précisons que pour les résultats obtenus grâce à cette méthode, nous avons appliqué le dictionnaire électronique morphologique des mots simples ainsi que le dictionnaire électronique existant des mots composés. Une autre possibilité serait d'appliquer uniquement le dictionnaire électronique des mots simples et ensuite comparer les résultats avec ceux que nous avons obtenus suite à l'application des deux dictionnaires. Nous n'avons pas effectué cette comparaison en raison du volume particulièrement élevé des candidats noms composés obtenus. La confrontation constituerait une étape trop importante.

L'avantage de la méthode de recensement semi-automatique est que nous trouvons les éléments lexicaux qui nous intéressent dans leur contexte naturel, c'est-à-dire dans des phrases qui ne sont pas fabriquées. Ceci nous permet de rendre compte de leur structure morphosyntaxique et de leur interprétation. Par ailleurs, l'utilisation des corpus pose le problème de la couverture de la langue. Tout corpus, aussi vaste soit-il, ne peut contenir toutes les structures possibles pour un terme, il ne peut donc pas rendre compte de l'ensemble de ses emplois ni représenter un phénomène linguistique dans toute sa complexité (M. Gross 1975).

Dans la liste finale, qui regroupe les séquences obtenues grâce aux deux méthodes de recensement, figurent 9 200 séquences, qui ont été soumises à des tests (ou critères) de délimitation des noms composés, afin d'être finalement qualifiées de noms composés de structure $N(E + DET:G)N:G$. Notons ici que le type des séquences recensées à partir des dictionnaires et des grammaires d'usage se différencie du type des séquences extraites à partir du corpus numérique. Dans le premier cas, nous avons surtout affaire à des noms composés lexicalisés, à des termes de spécialité et à des entités nommées désignant notamment des organisations et des lieux. En revanche, dans le deuxième cas, il s'agit majoritairement de syntagmes nominaux semi-figés (ou productifs) et d'entités nommées appartenant à toutes les catégories sémantiques.

Rappelons qu'à ces entrées s'ajoutent les 1 500 entrées retenues du DELAC-NNG actuel (cf. II, 4.2.1) et les 4 200 entrées obtenues des dictionnaires et des grammaires d'usage, également soumises à nos critères de délimitation des noms composés (cf. I, 2.2.1-2.2.11).

Ces réserves étant faites, il nous semble légitime d'estimer qu'un inventaire d'environ 11 500 formes de structure $N(E + DET:G)N:G$, construit à partir à la fois de sources éditées et de corpus numériques, constitue un échantillon représentatif permettant d'étudier les propriétés spécifiques du phénomène de la composition nominale en grec moderne.

Nous présentons par la suite notre proposition de classification des $N(E + DET:G)N:G$ en fonction des différentes sous-catégories identifiées lors de l'étape du recensement.

4.2.4 Proposition de sous-catégorisation des $N(E + DET:G)N:G$

Dans le cadre de cette étude, nous proposons une classification selon laquelle, les noms composés $N(E + DET:G)N:G$ sont répartis en cinq sous-catégories en fonction des propriétés spécifiques que présente chacune d'entre elles. Cette sous-catégorisation régit les étapes du recensement (cf. II, 4.2.1-4.2.3), de l'étude de la structure lexicale interne (cf. III, 6) et de la représentation formalisée des données en vue de leur reconnaissance lexicale automatique dans les textes (cf. V, 11.3). Cette classification s'applique à la fois aux noms composés binaires et aux noms composés complexes (ou surcomposés) (cf. III, 6.2 et 6.3). Les sous-catégories comprennent :

- les noms composés entièrement figés (e.g. φακός επαφής/lentille de contact) (cf. I, 2.3.1 et III, 6.4). Nous les distinguons davantage entre noms prédicatifs (e.g. (πάσχω

από/souffrir de) *σύνδρομο στέρησης/syndrome de sevrage*) et noms non-prédicatifs (e.g. *ζώνη ασφαλείας/ceinture de sécurité*) (cf. IV, 9) ;

- les syntagmes nominaux semi-figés (ou productifs) (e.g. *σχολή (διακοσμητικής + καλών τεχνών + τουριστικών επαγγελμαμάτων + μηχανολόγων + ...)/école* (de décoration + de beaux arts + d'hôtellerie + d'ingénieurs + ...) (cf. I, 2.3.2 et III, 6.5). Ils peuvent eux-aussi se répartir en prédicatifs (e.g. *έκθεση (ζωγραφικής + φωτογραφίας + βιβλίου +...)/exposition* (de peinture + de photographie + de livres +...)) ou non-prédicatifs (e.g. *κέντρο (έρευνας + τεκμηρίωσης + ιστορίας +...)/centre* (de recherche + de documentaion + d'histoire +...)). Notons que les syntagmes nominaux semi-figés (ou productifs) peuvent être déclencheurs (*i.e.* preuves internes ou externes) d'une entité nommée (cf. C. Martineau *et al.* 2007) (e.g. *ινστιτούτο (αισθητικής + ομορφιάς +...)* EN_lieu/institut (d'esthétique + de beauté +...) EN_lieu) ou pas (e.g. *άδεια (μητρότητας + κύησης + ασθενείας + ...)/congé* (de maternité + d'accouchement + de maladie + ...)) ;
- les entités nommées d'« organisation » (e.g. *Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης/Université Aristote de Thessalonique*), de « lieu » (e.g. *Γη του Πυρός/Terre de Feu*), d'« œuvre » (e.g. *Ερμής του Πραξιτέλη/Hermès de Praxitèle*) etc. Précisons que, dans le cadre de la présente recherche, nous ne cherchons pas à effectuer une étude complète, mais seulement une première classification des entités nommées (cf. III, 6.6) ;
- les termes de spécialité (e.g. *φάσμα μάζας/spectre de masse*). Ils se subdivisent, à leur tour, en fonction du domaine de spécialité auquel ils appartiennent (cf. III, 6.7) ;
- les syntagmes nominaux figés qui font partie d'une phrase simple figée mais qui ont un emploi autonome en dehors de celle-ci (e.g. *βάπτισμα του πυρός/baptême du feu*) (cf. III, 6.8).

Du point de vue linguistique, cette classification permet de rendre compte de certains phénomènes qui sont spécifiques à chaque sous-catégorie (par exemple, absence systématique de déterminant dans les syntagmes nominaux semi-figés (ou productifs), possibilité fréquente de substitution de la forme complète par la forme abrégée dans les entités nommées d'« organisation »). Elle permet également une meilleure organisation des entrées du DELACF. Du point de vue applicatif, la possibilité d'adaptation des données lexicales à des

applications spécifiques peut être facilitée ; par exemple, les sous-catégories du lexique peuvent être extraites plus facilement en fonction de l'application envisagée.

Chapitre 5. La variation

5.1 Introduction

L'étape du recensement nous a permis de rendre compte du phénomène de la variation au sein des $N(E+DET:G)N:G$. Les expériences effectuées sur des corpus de textes bruts en grec moderne ont montré que les variantes (lexicales, orthographiques, graphiques, phonologiques, syntaxiques) sont nombreuses dans le cas des $N(E+DET:G)N:G$. Dans certains cas, elles sont dues à la présence de formes de la langue savante (ou « katharevousa ») en grec moderne (cf. T. Kyriacopoulou 2003). Cependant, malgré leur présence importante en grec moderne, les variantes sont loin d'être systématiquement enregistrées dans les dictionnaires et les grammaires d'usage.

Dans ce chapitre, nous examinons les variantes lexicales, orthographiques, graphiques, phonologiques et syntaxiques des $N(E+DET:G)N:G$, attestées dans notre corpus (cf. II, 4.2.1-4.2.3). Notons que les variantes compliquent souvent la représentation formalisée des noms composés. Les solutions mises en œuvre ici pour faire face à ces problèmes sont exposées tout au long de ce chapitre.

Notre étude vise aux applications dans lesquelles le niveau de couverture lexicale de textes traités doit être très large. En l'absence d'applications concrètes, notre but est de fournir des données exhaustives – dans la mesure du possible – et adaptables à tout type d'application le plus facilement possible. « Rien n'indique *a priori* dans quels domaines se situent les textes susceptibles d'être traités par les applications informatiques ; de plus, notre étude vise aux applications dans lesquelles nous n'imposons aucune restriction de vocabulaire aux textes traités, qui sont entièrement « libres », c'est-à-dire qui obéissent aux seules contraintes de la langue elle-même. Ceci nous oblige à prendre en compte toutes les formes variées attestées dans les corpus textuels » (S. Voyatzi, 2006 : 133). Cependant, la délimitation entre formes variées acceptables et formes variées inacceptables n'est pas toujours très nette. « M. Mathieu-Colas (1990 : 107) résume la problématique concernant l'acceptabilité des formes variées ainsi : « En cas de doute, faut-il admettre assez largement les variantes (au risque d'inclure des formes périmées) ou restreindre la sélection (au risque d'être incomplets) ? ». Nous y répondons par la simple affirmation que les seules références pour l'acceptabilité des variantes sont les corpus de textes et les locuteurs natifs » (S. Voyatzi, 2006 : 133).

En ce qui concerne les noms composés grecs de structure $N(E+DET:G)N:G$, nous avons distingué les cinq types de variantes suivants :

- les variantes lexicales,
- les variantes orthographiques,
- la possibilité de substitution avec des formes abrégées,
- les variantes phonologiques,
- les variantes syntaxiques⁶⁹.

Pour faciliter la lecture de ce chapitre, nous étudions ces types de variantes premièrement au sein des noms composés binaires et ensuite au sein des noms composés complexes (ou surcomposés). Ce choix est d'autant plus motivé par le fait que dans le cas des noms composés complexes (ou surcomposés) le deuxième composant nominal est lui-aussi composé d'un adjectif et d'un nom ; les deux constituants forment ensemble une seule unité lexicale susceptible de différents types de variation. De manière générale, ces variantes affectent toutes les sous-catégories identifiées (cf. I, 2.3.1-2.3.2 et III, 6.2-6.8).

⁶⁹ Dans le cadre de cette étude, nous ne nous intéressons pas aux variantes référentielles, comme par exemple :

βασιλιάς των ζώων/roi des animaux
βασιλιάς των λουλουδιών/roi des fleurs
άστρο της ημέρας/étoile du jour
άστρο της νύχτας/étoile de la nuit

Ce sont des variantes référentielles de dénomination (familiale, connotée etc.) pour désigner respectivement le 'lion', la 'rose', le 'soleil' et la 'lune'.

Il s'agit d'un phénomène fréquent également dans le cadre des entités nommées. Par exemple :

άξονας του κακού/axe du mal

est utilisé pour désigner l'Union Soviétique.

« On remarque cependant la même variation au niveau des entités nommées qu'au niveau de termes. Il existe de multiples façons de désigner une personne ou un objet, il n'y a pas de nom unique et inévitable (cf. l'exemple de Frege autour de *l'étoile du soir* et *l'étoile du matin* qui désignent toutes deux Vénus...) » (T. Poibeau 2005).

A titre indicatif, citons ici quelques exemples de variantes référentielles pour certains emprunts ou termes de spécialité. Les formes :

χώρος στάθμευσης/aire de stationnement
αποσβεστήρας κραδασμών/amortisseur
νόσος του Μπάζεντοφ/maladie de Basedow

sont les variantes référentielles respectives pour « *πάρκινγκ*/aire de stationnement », « *αμορτισέρ*/amortisseur » et « *υπερθυρεοειδισμός*/hyperthyroïdie ».

Ces variantes figurent dans notre dictionnaire électronique morphologique et dans les tables du lexique-grammaire, mais elles ne sont pas associées à leur référent. Notons qu'on trouve presque toujours leur référent dans les dictionnaires d'usage. Leur étude systématique, qui nécessite une étude sémantique, notionnelle et ontologique, reste à faire.

5.2 Les noms composés binaires

5.2.1 Variantes lexicales

Les variantes lexicales sont susceptibles d'affecter les deux constituants « pleins » des noms composés binaires de structure $N (E+DET:G) N:G$ du grec moderne, à savoir le nom-tête et le deuxième composant nominal. Nous les présentons de manière détaillée dans ce qui suit.

5.2.1.1 Possibilité de substitution des formes savantes par des formes démotiques

i) Variantes du nom-tête

Nous avons observé certains cas de variation du nom-tête, où il y aurait la possibilité de substitution du nom-tête *savant* par un autre nom *démotique*, sans changement de sens. Ce type de substitution est propre à la langue grecque. Examinons les exemples suivants :

<i>κυτίο παραπόνων</i>	(variante savante)
<i>κουτί παραπόνων</i>	(variante démotique)
(boîte à réclamations)	
<i>κάλαθος (E + των) αχρήστων</i>	(variante savante)
<i>καλάθι (E + των) αχρήστων</i>	(variante démotique)
(corbeille à papier)	
<i>μάρτυς κατηγορίας</i>	(variante savante)
<i>μάρτυρας κατηγορίας</i>	(variante démotique)
(témoin à charge)	
<i>κοιλιάς του κλαυθμόνος</i>	(variante savante)
<i>κοιλιάδα του κλαυθμόνος</i>	(variante démotique)
vallée la- <i>Gms</i> lamentation- <i>Gms</i>	
(enfer)	

Dans les exemples précités, les deux formes variées des noms composés sont sémantiquement équivalentes. La possibilité de substitution des formes *savantes* par des formes *démotiques* n'est pas un phénomène spécifique des noms composés ; il s'observe au niveau des mots simples aussi. Ainsi, en ce qui concerne les deux formes variées *μάρτυς*/témoin et *μάρτυρας*/témoin, elles sont accompagnées d'un code flexionnel⁷⁰ par lequel elles sont liées

⁷⁰ Pour l'explicitation des codes flexionnels, cf. V, 11.3.1.1-11.3.1.3.

entre elles et représentées par la même forme canonique dans le dictionnaire électronique morphologique des mots simples (cf. V, 11.3.1). Nous avons donc gardé ce même code flexionnel dans le dictionnaire électronique morphologique des noms composés : par conséquent, les deux formes composées sont, elles aussi, reliées entre elles par la même forme canonique. A titre indicatif, citons l'extrait correspondant du dictionnaire électronique morphologique des formes fléchies⁷¹ :

μάρτυρας κατηγορίας, .N+NNG:Nms:Nfs
 μάρτυς κατηγορίας, μάρτυρας κατηγορίας.N+NNG:Nms:Vms:Nfs:Vfs
 μάρτυρα κατηγορίας, μάρτυρας κατηγορίας.N+NNG:Gms:Ams:Vms:Afs:Vfs
 μάρτυρος κατηγορίας, μάρτυρας κατηγορίας.N+NNG:Gms:Gfs
 μάρτυρες κατηγορίας, μάρτυρας κατηγορίας.N+NNG:Nmp:Vmp:Nfp:Afp:Vfp
 μαρτύρων κατηγορίας, μάρτυρας κατηγορίας.N+NNG:Gmp:Gfp

En ce qui concerne les formes savantes *κυτίο*/boîte, *κάλαθος*/corbeille et *κοιλιάς*/vallée, leur emploi autonome en dehors des noms composés *κυτίο παραπόνων*/boîte à réclamations, *κάλαθος (Ε + των) αχρήστων*/corbeille à papier et *κοιλιάς του κλαυθμώνος*/enfer est très limité en grec moderne. Pour cette raison, elles ne sont pas reliées aux formes démotiques respectives *κουτί*/boîte, *καλάθι*/corbeille et *κοιλάδα*/vallée dans le dictionnaire électronique morphologique des mots simples. Les noms composés dont les noms têtes sont *κουτί*/boîte, *καλάθι*/corbeille et *κοιλάδα*/vallée constituent également des entrées indépendantes dans le dictionnaire électronique morphologique des noms composés.

A l'opposé du phénomène de la possibilité de substitution des formes *savantes* par des formes *démotiques* pour le nom-tête se situe le phénomène de l'impossibilité de substitution de la forme savante du nom-tête, recensée pour un nom composé *N (Ε+DET:G) N:G*, par une forme démotique sémantiquement équivalente. A titre indicatif, citons l'exemple suivant :

γυνή της απωλείας (variante savante)
 (femme corrompue)

mais :

**γυναίκα της απωλείας* (variante démotique)
 (femme corrompue)

⁷¹ Sur les dictionnaires électroniques morphologiques cf. V, 11.3.1.

Dans nos données lexicales (cf. II, 4.2.1-4.2.3), le phénomène de l'impossibilité de substitution du nom-tête est peu fréquent. Nous nous en servons simplement afin d'identifier le caractère figé (ou afin de mesurer le degré de figement) des noms composés étudiés ici (cf. I, 2.2.6, 2.2.9).

ii) Variantes du deuxième composant nominal

Ce type de substitution s'observe également pour le deuxième composant nominal. On distingue les phénomènes suivants :

a) Variantes flexionnelles

Elles concernent les différentes marques flexionnelles du génitif et, éventuellement, le déplacement de l'accent lors de la flexion en grec savant (ou katharevoussa) par rapport à la flexion en grec populaire (ou démotique) (cf. V, 11.3.1.3).

Examinons les exemples suivants :

<i>αντιρρησίας συνειδήσεως</i>	(variante savante)
<i>αντιρρησίας συνείδησης</i>	(variante démotique)
(objecteur de conscience)	
<i>διαφορά φάσεως</i>	(variante savante)
<i>διαφορά φάσης</i>	(variante démotique)
(décalage de phase)	
<i>ένταλμα συλλήψεως</i>	(variante savante)
<i>ένταλμα σύλληψης</i>	(variante démotique)
(mandat d'arrêt)	
<i>αδυναμία χαρακτήρος</i>	(variante savante)
<i>αδυναμία χαρακτήρα</i>	(variante démotique)
(faiblesse de caractère)	
<i>άθλος του Ηρακλέους</i>	(variante savante)
<i>άθλος του Ηρακλή</i>	(variante démotique)
(travail d'Héraclès)	

<i>ορός (Ε + της) αλη<u>θ</u>είας</i>	(variante savante)
<i>ορός (Ε + της) αλη<u>θ</u>ειας</i>	(variante démotique)
(sérum de vérité)	

Les différentes positions des accents ainsi que les deux formes du génitif (savante et démotique) sont générées automatiquement à partir de la forme canonique du mot simple, grâce au code flexionnel qui lui est attribué et qui prévoit les deux flexions (savante et démotique). Dans ce cas, une seule entrée suffit dans le dictionnaire électronique morphologique des noms composés pour la génération automatique des deux formes du génitif⁷².

b) Variantes lexicales

Elles concernent la possibilité de substitution du mot savant par le mot démotique. A titre indicatif, citons les exemples suivants :

<i>αγιασμός των υ<u>δ</u>άτων</i>	(variante savante)
<i>αγιασμός των νε<u>ρ</u>ών</i>	(variante démotique)
(bénédiction des eaux)	
<i>παρανάλωμα του πυ<u>ρ</u>ός</i>	(variante savante)
<i>παρανάλωμα της φω<u>τ</u>ιάς</i>	(variante démotique)
(brûlure complète)	

Dans le cas où le phénomène de la substitution implique l'emploi d'un mot savant qui est synonyme d'un mot démotique (*αγιασμός των (υδάτων + νερών)*/bénédiction des eaux, *παρανάλωμα (του + της) (πυρός + φωτιάς)*/brûlure complète), nous avons opté pour la représentation des deux variantes par deux entrées autonomes dans le dictionnaire électronique morphologique des noms composés. Précisons que les deux variantes sont sémantiquement équivalentes.

Comme nous l'avons déjà signalé pour le nom-tête, nous avons remarqué l'impossibilité de substituer la forme savante du deuxième composant nominal par une forme démotique équivalente. A titre indicatif, citons les exemples suivants de variantes flexionnelles :

<i>(μήνας + σελήνη) του μέλι<u>τ</u>ος</i>	(variante savante)
--	--------------------

(lune de miel)

mais :

**(μήνας + σελήνη) του μελιού* (variante démotique)

(lune de miel)

δοχείο νυκτός (variante savante)

(pot de chambre)

mais :

**δοχείο νύχτας* (variante démotique)

(pot de chambre)

λάμπα θυέλλης (variante savante)

(lampe-tempête)

mais :

**λάμπα θύελλας* (variante démotique)

(lampe-tempête)

5.2.1.2 Possibilité de substitution des formes du grec par des emprunts

Il s'agit de la possibilité de substitution des formes grecques par des formes empruntées notamment au français, à l'anglais et à l'allemand et transcrites en alphabet grec (cf. A. Anastassiadis-Syméonidis 1986 : 141). Les dictionnaires usuels leur assignent la catégorie grammaticale de noms invariables.

i) Variantes du nom-tête

A titre d'illustration, citons les exemples suivants :

δοκιμασία κοπώσεως (forme grecque)

τεστ κοπώσεως (emprunt)

(épreuve d'anoxémie)

κινηματογράφος του δημιουργού (forme grecque)

⁷² Pour la génération automatique des formes fléchies à partir de la forme canonique, cf. V, 11.3.1.2.

<i>σινεμά του δημιουργού</i> (cinéma d'auteur)	(emprunt)
<i>φράγμα πυρός</i>	(forme grecque)
<i>μπαράζ πυρός</i> (écran coupe-feu)	(emprunt)
<i>κωμωδία ηθών</i>	(forme grecque)
<i>κομεντί ηθών</i> (comédie de mœurs)	(emprunt)
<i>σύμπλεγμα κατωτερότητας</i>	(forme grecque)
<i>κόμπλεξ κατωτερότητας</i> (complexe d'infériorité)	(emprunt)

Nous avons observé que ce phénomène de substitution de formes grecques par des emprunts n'est pas régulier. Il ne pourrait donc pas être généralisé, ni les termes de spécialité ni pour le vocabulaire général. A titre indicatif, citons les exemples suivants :

<i>σύμπλεγμα κατωτερότητας</i>	(forme grecque)
<i>κόμπλεξ κατωτερότητας</i> (complexe d'infériorité)	(emprunt)

mais :

<i>σύμπλεγμα της Ηλέκτρας</i>	(forme grecque)
<i>*κόμπλεξ της Ηλέκτρας</i> (complexe d'Electre)	(emprunt)

<i>φράγμα πυρός</i>	(forme grecque)
<i>μπαράζ πυρός</i> (écran coupe-feu)	(emprunt)

mais :

<i>φράγμα (Ε + του) ήχου</i>	(forme grecque)
<i>*μπαράζ (Ε + του) ήχου</i> (mur du son)	(emprunt)

ii) Variantes du deuxième composant nominal

D'après nos données lexicales (cf. II, 4.2.1-4.2.3), les variantes du deuxième composant nominal sont moins fréquentes :

γάντι της πυγμαχίας (forme grecque)

γάντι του μποξ (emprunt)

(gant de boxe)

σάκος της πυγμαχίας (forme grecque)

σάκος του μποξ (emprunt)

(poire de boxe)

αστέρι του κινηματογράφου (forme grecque)

αστέρι του σινεμά (emprunt)

(star du cinéma)

Les variantes présentées dans les exemples ci-dessus sont sémantiquement équivalentes et incluses dans le dictionnaire électronique des mots simples. Elles ne sont pas reliées aux formes grecques correspondantes. Nous avons procédé à l'illustration de chaque variante par une entrée indépendante dans le dictionnaire électronique morphologique des noms composés et dans les tables du lexique-grammaire.

5.2.1.3 Possibilité de substitution par des formes en latin

Nous avons observé des variantes en latin pour les deux noms composés suivants :

μέρα οργής

dies irae

αιτία πολέμου

casus belli

Les deux variantes sont sémantiquement équivalentes. Les variantes latines constituent des entrées autonomes.

5.2.1.4 Possibilité de substitution du nom-tête par son diminutif

Nous avons recensé un cas unique de substitution du substantif qui figure en tant que nom-tête par son diminutif :

(μάτι + ματάκι) της πόρτας
(judas de la porte)

Dans ce cas aussi nous avons représenté les deux formes variées par deux entrées dans le dictionnaire électronique morphologique et dans la table du lexique-grammaire.

5.2.1.5 Possibilité de substitution du deuxième composant nominal par une forme non-lexicale

Dans certains cas, le deuxième composant nominal peut être remplacé par un symbole. A titre indicatif, donnons l'exemple suivant :

ζώνη του (ευρώ + €)
(zone euro)

Le phénomène de la substitution du substantif par un symbole est presque marginal.

5.2.1.6 Variation de nombre

Selon M. Gross (1985 : 111), « le concept traditionnel de pluriel est une notion morphologique avant tout. Elle est simple et bien définie dans des exemples usuels comme

le-les, un-des, cheval-chevaux, mangerai-mangerons

où l'on peut parler d'articles et de noms pluriels, de verbe à la première personne du pluriel ».

De manière générale, la mise au pluriel des $N(E+DET:G)N:G$ est obtenue par la mise au pluriel du nom-tête uniquement. Le deuxième composant nominal reste invariable au génitif (singulier ou pluriel) lors de la flexion du nom composé. Ainsi, pour les exemples suivants :

πιάτο (E + της) ημέρας
plat (E + la- G_{fs}) journée- G_{fs}
(plat du jour)

επιταχυντής σωματιδίων
accélérateur particules- G_{fp}

(accélérateur de particules)

le nominatif pluriel sera respectivement :

πιάτα (E + της) ημέρας

plats-Nnp (E + la-Gfs) journée-Gfs

(plats du jour)

επιταχυντές σωματιδίων

accélérateurs-Nmp particules-Gfp

(accélérateurs de particules)

La forme suivante, qui impliquerait aussi la mise au pluriel du deuxième composant nominal aussi, n'est pas acceptable :

**πιάτα (E + των) ημερών*

**plats-Nnp (E + les-Gfp) journées-Gfp*

(*plats des jours)

Toutefois, la question de la mise au pluriel est plus complexe, et des problèmes de mise au pluriel se posent souvent. Nous avons observé que certains noms composés acceptent le deuxième composant nominal au génitif singulier ou au pluriel. A titre indicatif, citons les exemples suivants :

πεδίο (E + της + των) (μάχης + μαχών)

(champ de bataille)

διάδρομος (απογείωσης+ απογειώσεων)

(piste d'envol)

διάδρομος (προσγείωσης+ προσγειώσεων)

(piste d'atterrissage)

θέατρο (μαριονέτας + μαριονετών)

(théâtre de marionnettes)

Cependant, la mise au pluriel du nom-tête ne semble pas affecter le choix du nombre pour le deuxième complément nominal au génitif. Ainsi, la variante au singulier reste possible pour celui-ci. Dans ce cas, pour le $N (E+DET:G) N:G$:

πεδίο (E + της) μάχης
 champ (E + la- G_{fs}) bataille- G_{fs}
 (champ de batailles)

il faut envisager la mise au pluriel de trois façons différentes, comme en témoignent les exemples suivants :

- (i) *πεδίο (E + των) μαχών*
 champ (E + les- G_{fp}) batailles- G_{fp}
 (champ de batailles)
- (ii) *πεδία (E + της) μάχης*
 champs (E + la- G_{fs}) bataille- G_{fs}
 (champs de bataille)
- (iii) *πεδία (E + των) μαχών*
 champs (E + les- G_{fp}) batailles- G_{fp}
 (champs de batailles)

C'est un phénomène qui caractérise également les groupes nominaux libres de même structure. Par exemple, pour la mise au pluriel de :

η ζώνη του κοριτσιού
 la ceinture la- G_{ns} fille- G_{ns}
 (la ceinture de la fille)

il faut prendre en compte les trois possibilités suivantes :

- (i) *οι ζώνες του κοριτσιού*
 les ceintures la- G_{ns} fille- G_{ns}
 (les ceintures de la fille)
- (ii) *η ζώνη των κοριτσιών*
 la ceinture les- G_{np} filles- G_{np}

(la ceinture des filles)

- (iii) *οι ζώνες των κοριτσιών*
les ceintures les- G_{np} filles- G_{np}
(les ceintures des filles)

Dans le cas des noms composés, les données que nous avons rassemblées nous permettent d'estimer qu'il s'agit d'un phénomène qui n'est pas régulier et qui ne pourrait donc pas être généralisé. Ainsi, pour les exemples suivants :

πεδίο βολής
champ tir- G_{fs}
(parc d'artillerie)

πεδίο δράσης
champ action- G_{fs}
(champ d'action)

nous acceptons uniquement la mise au pluriel « régulière » des $N (E+DET:G) N:G$, à savoir :

πεδία βολής
champs tir- G_{fs}
(parc d'artillerie)

πεδία δράσης
champs action- G_{fs}
(champs d'action)

étant donné que les mises au pluriel suivantes :

(πεδίο + πεδία) βολών
(champ + champs) tirs- G_{fp}
(parcs d'artillerie)

(πεδίο + πεδία) δράσεων
(champ + champs) actions- G_{fp}
(champs d'action)

ne sont pas attestées dans le corpus étudié.

5.2.2 Variantes orthographiques

En grec moderne, certains mots simples présentent des variantes orthographiques. Ces mots simples peuvent constituer le nom-tête ou le deuxième composant nominal des noms composés de structure $N(E+DET:G)N:G$. Les deux variantes orthographiques restent également possibles au sein du $N(E+DET:G)N:G$. Ceci nous permet de parler ici de variantes orthographiques des noms composés de structure $N(E+DET:G)N:G$ du grec moderne. Notons que nous n'avons rencontré aucun cas de non-acceptabilité d'une des deux variantes orthographiques pour les cas recensés. Nous explicitons les variantes orthographiques dans ce qui suit.

i) Variantes du nom-tête

Comme nous l'avons dit, les variantes orthographiques peuvent affecter le nom-tête. A titre indicatif, citons les exemples suivants :

(αβγό + αυγό) ημέρας

(œuf du jour)

(δικλίδα + δικλείδα) ασφαλείας

(soupape de sûreté)

(διοξίδιο + διοξειδίο) του άνθρακα

(acide carbonique)

(κόλα + κόλλα) αναφοράς

(feuille de papier)

(εταιρία + εταιρεία) χαρτοφυλακίου

(société à portefeuille)

(παλικάρι + παλληκάρι) της φακής

(faux brave)

ii) Variantes du deuxième composant nominal

Dans nos données lexicales, nous avons rencontré beaucoup moins de variantes orthographiques au niveau du deuxième composant nominal. A titre d'illustration, citons l'exemple ci-dessous :

*μαγιά της (μπύρας + μύρας)*⁷³
(levure de bière)

En ce qui concerne la représentation formalisée des deux variantes orthographiques, nous proposons de les relier par la même forme canonique dans le dictionnaire électronique morphologique⁷⁴. Une des deux variantes orthographiques pourrait être choisie arbitrairement comme forme canonique. Seule la forme canonique devrait figurer dans le dictionnaire électronique morphologique (et dans les tables du lexique-grammaire).

5.2.3 Possibilité de substitution par des formes abrégées

Ce phénomène de variation n'affecte pas l'un des constituants « pleins » du nom composé (nom-tête ou deuxième complément nominal), mais le nom composé dans son ensemble. Il s'agit de la possibilité de substitution de la forme complète du nom composé par une ou plusieurs formes abrégées. A titre indicatif, citons les exemples suivants :

Προγνωστικά Ποδοσφαίρου (forme complète)
ΠΡΟ-ΠΟ (forme abrégée)
(pronostics de football)

Πρόεδρος (Ε + της) Βουλής (forme complète)
ΠΒ ou *Π.Β.* (formes abrégées)
(président de l'assemblée nationale)

Πρόεδρος Κυβερνήσεως (forme complète)
ΠΚ ou *Π.Κ.* (formes abrégées)
(président du gouvernement)

τεστ Παπανικολάου (forme complète)

⁷³ Chacune des variantes orthographiques de ce nom composé présente également une variante soudée : *μυρομαγιά*/levure de bière et *μυρομαγιά*/levure de bière respectivement. Pour les variantes soudées cf. II, 5.2.5.6.

⁷⁴ Pour l'instant, les deux formes variées figurent comme entrées indépendantes dans le dictionnaire électronique morphologique et dans les tables du lexique-grammaire. Le programme pour leur mise en relation dans le dictionnaire électronique morphologique est en construction (cf. V, 1.2.1).

τεστ παπ ou *παπ τεστ* (formes abrégées)
(pap test) (domaine : Médecine)

βιβλίο περιπτέρου (forme complète)

βίπερ (forme abrégée)

(livre de poche)

On peut noter que les formes abrégées peuvent comporter ou ne pas comporter des points (*ΠΒ* ou *Π.Β.*/président de l'assemblée nationale) ou le trait d'union (*ΠΠΟ-ΠΟ*/pronostics de football) en tant que séparateurs ou bien de constituer un mot simple (*βίπερ*/livre de poche).

En ce qui concerne leur représentation formalisée et reconnaissance automatique, les formes abrégées pourraient être reliées à la forme complète par la même forme canonique au moyen des tables du lexique-grammaire, comme ceci a été fait pour les variantes de type *AN* et les variantes soudées (cf. V, 11.3.3). Cependant, nous n'avons pas procédé à leur mise en relation, car il s'agit d'un phénomène de variation marginal dans nos données. Notons que la substitution de la forme complète par une forme abrégée est un phénomène de variation qui caractérise surtout les entités nommées (cf. III, 6.6).

5.2.4 Variantes phonologiques

Comme le note S. Voyatzi (2006 : 144) pour les adverbes, tout comme les autres catégories grammaticales de la langue grecque, les noms composés de structure *N (E+DET:G) N:G* sont eux-aussi susceptibles de variations phonologiques et donnent éventuellement lieu à des variantes graphiques. Dans la majorité des cas, les variantes phonologiques ne sont pas répertoriées de façon systématique dans les dictionnaires d'usage. Notre objectif étant la construction d'un dictionnaire électronique destiné à l'analyse automatique des textes, nous devons prendre en compte et représenter de manière explicite les variantes phonologiques aussi.

En nous basant sur nos données lexicales, nous avons regroupé les variantes phonologiques des noms composés de structure *N (E+DET:G) N:G* en deux grandes catégories :

i) Variantes dues aux transformations de voyelles

a) l'aphérèse

Le phénomène d'aphérèse concerne un mot, dont la voyelle initiale est optionnelle, comme le montre l'exemple suivant :

ημέρα + μέρα

(jour)

L'emploi des deux formes variées n'est pas toujours possible. Nous avons observé que, dans la majorité des cas, le mot *ημέρα*/jour ne peut pas être substitué par la variante *μέρα*/jour :

*κρέμα (ημέρας + *μέρας)*

(crème de jour)

*(αβγό + αυγό) (ημέρας + *μέρας)*

(œuf du jour)

Cependant, il ne s'agit pas d'un phénomène régulier :

φως της (ημέρας + μέρας)

(lumière du jour)

En revanche, dans le cas des noms composés où le mot *ημέρα*/jour figure en tant que nom-tête, sa substitution par la variante *μέρα*/jour est possible :

(ημέρα + μέρα) οργής

(jour de colère)

b) le changement de voyelle

Nous avons observé le phénomène phonologique du changement de /o/ en /u/ dans l'exemple suivant :

(λωρίδα + λουρίδα) κυκλοφορίας⁷⁵

(voie de circulation)

Dans l'exemple suivant, on observe le même phénomène phonologique ; cette fois c'est le /e/ qui change en /io/ :

⁷⁵ Selon A. Anastassiadis-Symeonidis et A. Fliatouras (2003), il ne s'agit pas ici uniquement d'une variante phonologique, mais aussi d'un phénomène phonologique qui témoigne de l'alternance d'une variante *savante*

(γεφύρι + γιοφύρι) της Άρτας⁷⁶
(pont d'Arta)

Rappelons que ce nom composé peut être soit une entité nommée de lieu, qui désigne un pont qui existe dans la ville grecque Arta, soit un nom composé commun, qui, selon un emploi métaphorique, désigne une œuvre qui ne peut pas être achevée (cf. 2.2.11). Précisons ici qu'il figure dans notre dictionnaire électronique et dans les tables du lexique-grammaire uniquement en tant que nom composé commun et dans la liste des entités nommées en tant qu'entité nommée de lieu.

c) la syncope

Le phénomène de syncope affecte un mot dont la voyelle, qui se trouve entre deux consonnes, est optionnelle. Il s'observe au niveau de mots simples, comme par exemple :

(κορυφή + κορυφή)
(sommets)

Mais, d'après nos données lexicales, il n'est pas autorisé dans les noms composés de structure $N(E+DET:G)N:G$ dont ce même mot est un de leurs constituants, comme en témoigne l'exemple suivant :

σύνοδος (κορυφής + *κορυφής)
(sommets)

ii) Variantes dues à la consonne épenthétique -v/n/

Nous avons observé ce phénomène uniquement dans l'exemple suivant :

(μήλον + μήλο) της έριδος
(pomme de discorde)

(*λωρίδα* κυκλοφορίας/voie de circulation) avec une variante populaire (*λουρίδα* κυκλοφορίας/voie de circulation). Nous traduisons le terme grec *λαϊκός* par le terme français *populaire*.

⁷⁶ Les deux variantes phonologiques ont également une variante permutée avec redoublement de l'article défini : της Άρτας το (γεφύρι + γιοφύρι)/la-Gfs Arta-Gfs le pont. Les variantes de ce type sont étudiées dans II, 5.2.5.4. Comme dans l'exemple précédent, selon A. Anastassiadis-Symeonidis et A. Fliatouras (2003), il ne s'agit pas ici uniquement d'une variante phonologique, mais aussi d'un phénomène phonologique qui témoigne de l'alternance d'une variante savante (γεφύρι της Άρτας/pont d'Arta) avec une variante populaire (γιοφύρι της Άρτας/pont d'Arta). Nous traduisons le terme grec *λαϊκός* par le terme français *populaire*.

Soulignons que le choix de la variante dépend largement du registre de la langue, du contexte et de l'énonciation (cf. K. Papanastasiou 2001).

Nous proposons de relier les variantes phonologiques et les formes pleines par la même forme canonique dans le dictionnaire électronique morphologique.

5.2.5 Variantes syntaxiques

Les variantes syntaxiques sont nombreuses dans les noms composés de structure $N(E+DET:G)N:G$. Elles affectent tous les constituants, c'est-à-dire le nom-tête, le déterminant et le deuxième composant nominal. Nous avons repéré divers types de variantes telles que des variantes de structure (par exemple : des formes réduites), des variantes de type AN et des variantes soudées. Nous avons effectué la représentation formalisée de toutes les variantes syntaxiques en les reliant à la forme canonique correspondante dans des tables du lexique-grammaire (cf. V, 11.2.1). Nous les explicitons dans ce qui suit.

5.2.5.1 La réduction du déterminant

Comme nous le signalons dans III, 7.2.1.1.1 et 7.2.1.2.3, les formes sans déterminant constituent souvent des variantes réduites des formes avec déterminant. Les déterminants des noms composés de structure $N(E+DET:G)N:G$ peuvent donc souvent être optionnels. La possibilité de substitution la plus fréquente que nous avons observée est celle du déterminant défini générique par le déterminant zéro, indicatif également de généricité. A titre d'illustration, citons les exemples suivants :

*ανακοπή (E + της) καρδιάς*⁷⁷

arrêt (E + le-*Gfs*) cœur-*Gfs*

(arrêt du cœur)

κλάσμα (E + του) δευτερολέπτου

fraction (E + la-*Gns*) seconde-*Gns*

(fraction de seconde)

νήμα (E + της) στάθμης

fil (E + le-*Gfs*) niveau-*Gfs*

(niveau à plomb)

⁷⁷ Cette forme présente également une variante réduite de type $N(E+DET:G)N:G = N$. Les variantes de ce type sont étudiées dans III, 7.2.5.3.

πιάτο (E + της) ημέρας
 plat (E + la-*Gfs*) journée-*Gfs*
 (plat du jour)

γλυκό (E + του) κουταλιού
 gâteau (E + la-*Gns*) cuiller-*Gns*
 (gâteau de cuillère)

διάσπαση (E + του) ατόμου
 fission (E + le-*Gns*) atome-*Gns*
 (fission de l'atome)

Cependant, la réduction de déterminant des noms composés de structure $N(E+DET:G)N:G$ peut affecter l'article indéfini aussi. A titre d'illustration, citons l'exemple suivant :

πυκνότητα (E + ενός) σώματος
 densité (E + un-*Gns*) corps-*Gns*
 (densité d'un corps)

La présence du déterminant indéfini en général et sa substitution par le déterminant zéro en particulier sont rares dans les $N(E+DET:G)N:G$. Il s'agit toujours de l'emploi générique, emploi que prend également le déterminant zéro.

5.2.5.2 La réduction du nom-tête

Un autre type de variante de structure concerne la réduction du nom-tête. Il s'agit d'un phénomène marginal dans les $N(E+DET:G)N:G$. A titre indicatif, donnons les exemples suivants :

(E + εισιτήριο) διαρκείας
 (carte d'abonnement)

(E + τμήμα) μεταγωγών
 (section transfèrement)

(E + τμήμα) αλλοδαπών
 (section étranger)

Les deux variantes sont sémantiquement équivalentes. Les formes réduites restent invariables au génitif, mais elles peuvent être employées à la place des formes complètes (cf. aussi I, 2.6).

Par exemple :

Το (Ε + εισιτήριο) διαρκείας δεν ισχύει για τη διαδρομή αυτή

(La carte d'abonnement n'est pas valide pour ce trajet)

Χθες αγόρασα (Ε + εισιτήριο) διαρκείας

(Hier j'ai acheté une carte d'abonnement)

5.2.5.3 La réduction du deuxième composant nominal

La réduction peut également affecter le deuxième composant nominal au génitif. Comme le note G. Gross (1996a : 47), « pour les noms composés à nom-tête, on peut observer des effacements », soit hors contexte soit en contexte situationnel. En guise d'illustration, citons les exemples suivants :

αλάτι (Ε + μαγειρικής)

(sel de cuisine)

δίπλωμα (Ε + οδήγησης)

(permis de conduire)

αγιασμός (Ε + των υδάτων)

(bénédiction des eaux)

γέεννα (Ε + του πυρός)

(géhenne)

επόπτης (Ε + γραμμών)

(arbitre de touche)

άρμα (Ε + μάχης)

(char de combat)

δίσκος (Ε + βινυλίου)

(disque vinyle)

La réduction du deuxième composant nominal est assez fréquente dans les $N(E+DET:G)N:G$. G. Gross (1996a : 47) remarque que « la possibilité d’effacement est un indice de moindre figement et en tout cas la preuve qu’il y a un substantif-tête que l’on interprète sans hésitation dans une situation donnée ». Cependant, dans la majorité des cas, les formes réduites posent des problèmes d’ambiguïté tant au niveau de la description linguistique qu’au niveau du traitement automatique. Par exemple, la forme réduite *δίπλωμα*/permis peut correspondre à plusieurs formes complètes :

δίπλωμα (οδήγησης + ευρεσιτεχνίας + σπουδών + ...)
 (permis de conduire + patente + diplôme d’études + ...)

Les cas où la forme réduite ne pose pas de problème d’ambiguïté sont très rares. A titre indicatif, citons l’exemple suivant :

γέεννα (Ε + του πυρός)
 (géhénne)

Dans ce cas, nous considérons que les deux formes variées sont sémantiquement équivalentes. Nous avons effectué la mise en relation des deux variantes dans les tables du lexique-grammaire, tout en notant que dans le cas de la variante réduite on a affaire à un nom ambigu. Précisons cependant que cette association n’est pas exploitable au niveau de la phrase, mais uniquement au niveau textuel.

5.2.5.4 La permutation

Comme nous l’avons vu (cf. II, 4.2.1), il existe des séquences où le composant nominal au génitif – qui est canoniquement postposé – est obligatoirement antéposé par rapport au nom-tête. Les $N(E+DET:G)N:G$ qui possèdent les deux variantes permutées, à savoir celle avec le composant nominal au génitif postposé ou antéposé, sont très rares. En guise d’illustration, citons les exemples suivants :

το μαλλί της γριάς + της γριάς το μαλλί
 (barbe à papa)

το (γεφύρι + γιοφύρι) της Άρτας + της Άρτας το (γεφύρι + γιοφύρι)
 (pont d’Arta)

Comme nous l'avons déjà signalé, en cas de permutation, le nom-tête reçoit la détermination définitive si le nom au génitif est déterminé par l'article défini. Il s'agit du phénomène de redoublement de l'article défini, propre à la langue grecque (I. Tsamadou 1984). Nous n'avons pas procédé à la représentation formalisée de ce type de variantes, car, d'après nos données lexicales, il s'agit d'un phénomène marginal.

5.2.5.5 Les variantes de type AN

Nous avons observé que l'adjectivation du deuxième complément nominal au génitif est un phénomène assez courant dans les noms composés grecs de structure $N (E+DET:G) N:G$. Le complément du nom au génitif accepte une paraphrase en A, où A est un adjectif morphologiquement associé au $N:G$, noté $N:G-a$. Les deux variantes sont synonymes. Cette relation peut être représentée par la construction générale suivante :

$$N (E+DET:G) N:G = N:G-a N$$

A titre indicatif, citons les exemples suivants :

αλάτι μαγειρικής = μαγειρικό αλάτι

(sel de cuisine)

βαλβίδα ασφαλείας = ασφαλιστική βαλβίδα

(soupape de sûreté)

διάσειση (Ε + του) εγκεφάλου = εγκεφαλική διάσειση

(commotion cérébrale)

δίκαιο των Ενοχών = ενοχικό δίκαιο

(droit des obligations)

δομή βάθους = βαθιά δομή

(structure profonde)

(εταιρεία + εταιρία) μεταφορών = μεταφορική (εταιρεία + εταιρία)

(entreprise de transport)

λογαριασμός (τράπεζας + τραπεζής) = τραπεζικός λογαριασμός

(compte bancaire)

πύελος νεφρού = νεφρική πύελος
(bassinets rénal)

σύμπλεγμα του Οιδίποδα = οιδιπόδειο σύμπλεγμα
(complexe d'Œdipe)

σύνδρομο της Κατοχής = κατοχικό σύνδρομο
(syndrome de l'occupation)

τσάι του βουνού = βουνίσιο τσάι
(thé des montagnes)

σάκος (Ε + της) πυγμαχίας = πυγμαχικός σάκος
(poire de boxe)

Parfois, il existe plusieurs *N:G-a* pour le même *N:G*. Lorsque ceux-ci sont sémantiquement équivalents, l'emploi de tous les *N:G-a* est possible lors de la formation du *N:G-a N* :

(αβγό + αυγό) του Πάσχα = (πασχαλιάτικο + πασχαλινό) (αβγό + αυγό)
(œuf de Pâques)

Quand ils ne sont pas sémantiquement équivalents, ils donnent lieu à des *N:G-a N* distincts et il n'y a pas de possibilité de substitution au sein du même *N:G-a N* :

*δικαιώματα του ανθρώπου = (ανθρώπινα + *ανθρωπιστικά δικαιώματα)*
(droits de l'homme)

*επιστήμες του ανθρώπου = (ανθρωπιστικές + *ανθρώπινες) επιστήμες*
(sciences humaines)

Des points de vue syntaxique et morphologique, l'adjectivation du *N:G* est postérieure à la formation du groupe nominal en *N:G*. La relation entre le complément du nom et l'adjectif morphologiquement associé à celui-ci, notée dans la présente étude *N:G = N:G-a*, n'est pas systématique et doit être examinée au cas par cas (cf. aussi A. Monceaux 1993 : 378). D'ailleurs, le groupe nominal formé par paraphrase du *N:G* en *N:G-a* ne correspond pas forcément au sens du *N (E+DET:G) N:G* :

έργο του σατανά ≠ σατανικό έργο

(œuvre du diable ≠ œuvre diabolique)

Nous avons enregistré systématiquement la relation $N(E+DET:G) N:G = N:G-a N$ lors du recensement de nos données (cf. II, 4.2.1-4.2.3). Nous avons effectué la mise en relation des deux variantes dans les tables du lexique-grammaire (cf. V, 11.2.1, 11.3.3 et 11.3.3.3)⁷⁸.

5.2.5.6 Les variantes soudées

Nous avons observé également qu'à un nom composé de structure $N(E+DET:G) N:G$ peut correspondre fréquemment un nom simple : ce nom simple constitue un nom composé « soudé ». Dans ce cas, le deuxième composant nominal du $N(E+DET:G) N:G$ devient le premier composant nominal du nom soudé. Le genre grammatical du nom-tête est le même que celui du nom soudé. A noter que ce type de variation n'est pas systématique et doit être étudié au cas par cas. A titre indicatif, citons les exemples suivants :

ανίχνευση του πυρός = πυρανίχνευση

(détection du feu)

σίδερο ατμού = ατμοσίδερο

(fer à repasser vapeur)

βόμβα υδρογόνου = υδρογονοβόμβα

(bombe à hydrogène)

δωμάτιο ύπνου = υπνοδωμάτιο

(chambre à coucher)

⁷⁸ Nous avons également observé des cas de synonymie entre un nom composé de structure $N(E+DET:G) N:G$ et un nom composé de structure AN où le nom-tête est le même, mais le A du AN n'est pas morphologiquement associé au $N:G$ du $N(E+DET:G) N:G$. A titre indicatif, citons les exemples suivants :

ποδήλατο γυμναστικής = (στατικό + σταθερό) ποδήλατο

(vélo statique)

μήκος τόπου = γεωγραφικό μήκος

(longitude)

De plus, dans certains cas, les $N(E+DET:G) N:G$ ont des synonymes de type AN , mais sans qu'il y ait un lien morphologique quelconque entre les constituants des noms composés des deux structures. Par exemple :

εσχάτη των ποινών = θανατική ποινή

(peine capitale)

Dans ce cas, nous considérons que nous avons affaire à des entrées indépendantes. Nous n'effectuons aucune mise en relation entre elles dans les tables du lexique-grammaire. Seules les entrées $N(E+DET:G) N:G$ figurent dans notre dictionnaire électronique des noms composés.

κατοχή όπλου = οπλοκατοχή
(détention d'armes)

κηλίδα πετρελαίου = πετρελαιοκηλίδα
(marée noire)

πηγή πετρελαίου = πετρελαιοπηγή
(source de pétrole)

πτώση βροχής = βροχόπτωση
(précipitation atmosphérique)

πτώση χιονιού = χιονόπτωση
(chute de neige)

σάκος ύπνου = υπνόσακος
(sac de couchage)

ψάρι του αφρού = αφρόψαρο
(poisson d'écume)

Dans le cadre de la substitution des noms composés de structure $N (E+DET:G) N:G$ par des noms composés soudés, nous avons observé certains phénomènes (cas particuliers) qui méritent d'être signalés. Nous les explicitons dans ce qui suit :

- i) le $N:G$ démotique du $N (E+DET:G) N:G$ devient savant lors de la formation de la variante soudée :

λάδι (E + της) ελιάς = ελαιόλαδο
(huile d'olive)

πυρήνας της ελιάς = ελαιοπυρήνας
(noyau d'olive)

- ii) inversement, le $N:G$ savant du $N (E+DET:G) N:G$ devient démotique lors de la formation de la variante soudée :

θέατρο ανδρεικέλων = κουκλοθέατρο
(théâtre de marionnettes)

- iii) les deux constituants du $N (E+DET:G) N:G$ sont des mots grecs, alors que le nom composé soudé est un mot emprunté à une langue étrangère :

έλεγχος οينوπνεύματος = αλκοτέστ
(alcootest)

- iv) le nom composé soudé est formé du $N:G$ du $N (E+DET:G) N:G$ et d'un nom synonyme du nom-tête :

δίνη ανέμου = ανεμοστρόβιλος
(tourbillon)

στήλη πάγου = παγοκολόνα
(colonne de glace)

- v) le $N:G$ du $N (E+DET:G) N:G$ est lui-même un nom composé soudé. Lors de la formation du nom composé soudé synonyme du $N (E+DET:G) N:G$ on retrouve uniquement le premier composant du $N:G$:

πέδιλο παγοδρομίας = παγοπέδιλο
(patin)

- vi) l'orthographe du nom-tête du $N (E+DET:G) N:G$ change lors de la formation de la variante soudée :

λατρεία των ειδώλων = ειδωλολατρία
(paganisme)

Nous avons enregistré systématiquement ces variantes lors du recensement de nos données (cf. II, 4.2.1-4.2.3). Nous avons effectué la mise en relation des deux variantes dans les tables du lexique-grammaire (cf. V, 11.2.1, 11.3.3 et 11.3.3.3).

5.2.5.7 Les variantes en $N \text{ Prép } (E+DET) N$

Nous avons également remarqué la possibilité de substitution de la structure $N (E+DET:G) N:G$ par la structure $N \text{ Prép } (E+DET) N$. D'après nos données lexicales, il s'agit d'un phénomène de variation marginal. A titre indicatif, citons l'exemple suivant :

ποτήρι (E+του) κρασιού = ποτήρι για κρασί

(verre à vin)

A noter que les deux formes sont sémantiquement équivalentes uniquement quand elles désignent l'objet concret. Il n'y a pas de possibilité de substitution quand elles désignent le nom de mesure :

Η Μαρία ήπια δύο ποτήρια του κρασιού χυμό

** Η Μαρία ήπια δύο ποτήρια για κρασί χυμό*

(Maria a bu deux verres à vin de jus)

Nous n'avons pas procédé à la représentation formalisée de ce type de variantes en raison du nombre extrêmement limité des cas attestés dans nos données lexicales.

5.2.5.8 Les $N (E+Prép) (E+DET:G) N:G$

Enfin, nous avons observé des cas où les noms composés de structure $N (E+DET:G) N:G$ sont des variantes réduites d'une forme avec préposition de type $N (E+Prép) (E+DET:G) N:G$, comme par exemple :

κυρία (E + επί) των τιμών

(dame d'honneur)

Il s'agit également d'un phénomène de variation marginal ; pour cette raison, nous n'avons pas effectué la mise en relation des deux formes variées.

En guise de conclusion, signalons que les différents types de variantes affectent environ 40% des $N (E+DET:G) N:G$ étudiés. Notons cependant que ce pourcentage arriverait jusqu'à 80% si on prenait en compte certains types de variantes qui sont très courantes au sein de certaines catégories, comme par exemple la substitution par des formes abrégées, qui est un phénomène très courant dans le cadre des entités nommées.

5.3 Les noms composés complexes (ou surcomposés) de structure $N1 (E+DET:G) A:G N2:G$

Tout comme les noms composés binaires, les noms composés complexes (ou surcomposés) peuvent avoir différents types de variantes (lexicales, orthographiques, graphiques ou syntaxiques). Par exemple :

άδεια (άσκησης + ασκήσεως) επαγγέλματος
(autorisation d'exercice)

(ανοιχτή + ανοικτή) γραμμή επικοινωνίας
(ligne ouverte)

έγκαυμα (πρώτου + 1^{ov}) βαθμού
(brûlure de premier degré)

Comme il a été explicité dans III, 6.3, dans le cadre de cette étude, nous nous sommes limitée à l'étude des noms composés complexes (ou surcomposés) qui sont définis par la structure morphosyntaxique *NI (E+DET:G) A:G N2:G*. Ce choix est motivé par le fait qu'ils sont les plus nombreux parmi les différents types de noms composés complexes (ou surcomposés) recensés⁷⁹. Les sections qui suivent sont consacrées à la description de leur variation lexicale, orthographique, graphique et syntaxique. La plupart de ces phénomènes ont déjà été étudiés et des solutions ont été proposées pour les noms composés binaires (cf. II, 5.2). C'est pourquoi nous n'insistons pas sur leur description ; nous nous contentons de donner quelques exemples.

5.3.1 Variantes lexicales

5.3.1.1 Possibilité de substitution des formes savantes par des formes démotiques

i) Variantes de l'adjectif

On distingue les deux cas suivants :

a) Variantes flexionnelles

Elles concernent les différentes marques flexionnelles du génitif et le déplacement de l'accent lors de la flexion en grec *savant* par rapport à la flexion en grec *populaire*, comme par exemple :

κατάσταση εκτάκτου ανάγκης	(variante <i>savante</i>)
κατάσταση έκτακτης ανάγκης	(variante <i>démotique</i>)
(état d'urgence)	

b) Variantes lexicales

Il s'agit de la substitution d'un adjectif *savant* par un adjectif *démotique* faisant partie du même paradigme synonymique. Par exemple :

έργο μακράς πνοής	(variante <i>savante</i>)
έργο μεγάλης πνοής	(variante <i>démotique</i>)
(œuvre de longue haleine)	

ii) Variantes du A:G N2:G

Examinons l'exemple suivant :

όνειρο θερινής νυκτός ⁸⁰	(variante <i>savante</i>)
όνειρο καλοκαιρινής νύχτας	(variante <i>démotique</i>)
(songe d'une nuit d'été)	

Dans l'exemple précité, les deux formes variées sont sémantiquement équivalentes. Notons ici que sont uniquement permises les formes variées dans lesquelles les deux éléments (adjectif et nom) sont savants ou démotiques. En revanche, des formes variées qui contiennent des éléments savants et démotiques mélangés ne sont pas autorisées :

*όνειρο θερινής νύχτας	(savant + démotique)
*όνειρο καλοκαιρινής νυκτός	(démotique + savant)
(songe d'une nuit d'été)	

iii) Variantes du N2:G

On distingue les deux cas suivants :

a) Variantes flexionnelles

Elles concernent les différentes marques flexionnelles du génitif et le déplacement de l'accent lors de la flexion en grec *savant* par rapport à la flexion en grec *populaire*, comme par exemple :

αλεύρι ολικής (αλέσεως + άλεσης)

⁷⁹ L'étude de la variation au sein des autres types de noms composés complexes (ou surcomposés) identifiés reste à faire.

⁸⁰ Ce nom composé complexe (ou surcomposé) peut être soit une entité nommée d'« œuvre » soit un nom commun. Dans l'exemple, il est pris comme nom commun.

(farine complète)

αλιεία ανοιχτής (θαλάσσης + θάλασσας)

(grande pêche)

δρόμος διπλής (κατευθύνσεως + κατεύθυνσης)

(route à double voie)

δρόμος μονής (κατευθύνσεως + κατεύθυνσης)

(sens unique)

εμπόριο λευκής (σαρκός + σάρκας)

(traite des blanches)

b) Variantes lexicales

Il s'agit de la possibilité de substitution d'un nom *savant* par un nom *démotique* faisant partie du même paradigme synonymique. Par exemple :

άγγελος κακών (επών + ειδήσεων)

(porteur de mauvaises nouvelles)

5.3.1.2 Possibilité de substitution d'un mot par un autre mot du même registre

i) Variantes du nom-tête

Dans les exemples suivants, le nom-tête donne lieu à un paradigme synonymique restreint :

(περίοδος + εποχή) (Ε + των) ισχνών αγελάδων

(période néfaste)

(περίοδος + εποχή) (Ε + των) παχιών αγελάδων

(période faste)

(πολιτική + διπλωματία) ανοικτών θυρών

(politique de la porte ouverte)

ii) Variantes de l'adjectif

Dans l'exemple suivant, c'est l'adjectif qui donne lieu à un paradigme synonymique restreint :

γραμμή (διαρκούς + αιώνιου) χιονιού
(neiges éternelles)

iii) Variantes du deuxième composant nominal

Dans les exemples suivants, le deuxième complément nominal donne lieu à un paradigme synonymique restreint :

μέσα μαζικής (ενημέρωσης + επικοινωνίας)
moyens massive- G_{fs} (actualité- G_{fs} + communication- G_{fs})
(communications de masse)

σύμβαση αποικιακού (χαρακτήρα + τύπου)
contrat colonial- G_{ms} (caractère- G_{ms} + type- G_{ms})
(contrat de type colonial)

5.3.1.3 Possibilité de substitution du singulier par le pluriel

De manière générale, la mise au pluriel des $N (E+DET:G) N:G = NI (E+DET:G) A:G N2:G$ est obtenue par la mise au pluriel du nom-tête uniquement. L'adjectif et le deuxième composant nominal restent invariables au génitif (singulier ou pluriel) lors de la flexion du nom composé complexe (ou surcomposé). Ainsi, pour l'exemple suivant :

έγκλημα εσχάτης προδοσίας
crime suprême- G_{fs} trahison- G_{fs}
(crime de lèse-majesté)

le nominatif pluriel sera :

εγκλήματα εσχάτης προδοσίας
crimes suprême- G_{fp} trahison- G_{fp}
(crimes de lèse-majesté)

Cependant, concernant la mise au pluriel de certains noms composés complexes (ou surcomposés)⁸¹, nous avons observé qu'il existe deux formes variées sémantiquement équivalentes : la première avec le $A:G N2:G$ au singulier et la deuxième avec le $A:G N2:G$ au pluriel. A titre indicatif, citons les exemples suivants :

μέσο (μαζικής μεταφοράς + μαζικών μεταφορών)
(moyen de transport collectif)

ψάρι (του γλυκού νερού + των γλυκών νερών)
(poisson d'eau douce)

Dans les exemples ci-dessus, la mise au pluriel du nom-tête n'affecte pas le nombre du $A:G\ N2:G$, pour lesquels la variante au singulier reste possible :

μέσα (μαζικής μεταφοράς + μαζικών μεταφορών)
(moyens de transport collectif)

ψάρια (του γλυκού νερού + των γλυκών νερών)
(poissons d'eau douce)

5.3.2 Variantes orthographiques

A titre indicatif, citons quelques exemples de variation orthographique du nom-tête et de l'adjectif :

(εταιρεία + εταιρία) περιορισμένης ευθύνης
(société à responsabilité limitée)

επιχείρηση (μεικτής + μικτής) οικονομίας
(entreprise d'économie mixte)

5.3.3 Possibilité de substitution par des formes abrégées

La possibilité de substitution par des formes abrégées est un phénomène qui s'observe également au sein des $NI\ (E+DET:G)\ A:G\ N2:G$ du grec moderne, comme en témoignent les exemples suivants :

*αυτοκίνητο δημόσιας χρήσης*⁸² (forme complète)

ΔΧ ou *Δ.Χ.* (formes abrégées)

(véhicule de service public)

⁸¹ Pour les problèmes de mise au pluriel, cf. aussi II, 5.2.1.6.

⁸² Dans le cas de cet exemple, la variante abrégée affecte l'adjectif et le deuxième complément au génitif ($A:G\ N2:G$). Cependant, la forme abrégée peut remplacer la forme complète du nom composé complexe (ou surcomposé) dans son ensemble.

αυτοκίνητο ιδιωτικής χρήσης ⁸³	(forme complète)
ΙΧ ou Ι.Χ.	(formes abrégées)
(véhicule routier à usage privé)	
δελτίο αστυνομικής ταυτότητας	(forme complète)
ΔΑΤ ou Δ.Α.Τ.	(formes abrégées)
(carte d'identité)	
(εταιρεία + εταιρία) περιορισμένης ευθύνης	(forme complète)
ΕΠΕ ou Ε.Π.Ε.	(formes abrégées)
(société à responsabilité limitée)	
μέσα μαζικής (ενημέρωσης + επικοινωνίας)	(forme complète)
ΜΜΕ ou Μ.Μ.Ε.	(formes abrégées)
(communications de masse)	
μονάδα εντατικής θεραπείας	(forme complète)
ΜΕΘ ou Μ.Ε.Θ.	(formes abrégées)
(unité de soins intensifs)	
φόρος προστιθέμενης αξίας	(forme complète)
ΦΠΑ ou Φ.Π.Α.	(formes abrégées)
(taxe à la valeur ajoutée)	

5.3.3 Variantes syntaxiques

Dans le cadre des $N1 (E+DET:G) A:G N2:G$ du grec moderne, nous avons rencontré les cas de variation syntaxique suivants :

- i) Réduction du deuxième composant nominal et substantivation de l'adjectif, notée $N1 (E+DET:G) A:G N2:G = N1 (E+DET:G) A-n:G$. La variante réduite constitue donc un nom composé binaire de structure $N (E+DET:G) N:G$. A titre indicatif, citons les exemples suivants :

άγγελος κακών ($E + ειδήσεων + επών$)
(porteur de mauvaises nouvelles)

⁸³ Cf. *Supra*.

αποκόλληση (E + του) αμφιβληστροειδούς (E + χιτώννα)
(décollement de la rétine)

λειτουργία των προηγιασμένων (E + δώρων)
(messe des Présanctifiés)

- ii) Adjectivation du deuxième composant nominal $A:G$ $N2:G$, qui constitue lui-même un nom composé, par la formation d'un adjectif composé soudé sémantiquement équivalent au nom composé AN , notée $N1$ ($E+DET:G$) $A:G$ $N2:G$ = $\langle A:G$ $N2:G \rangle$ -a $N1$. Par exemple :

άνθρωπος κακής πίστεως = κακόπιστος άνθρωπος
(personne de mauvaise foi)

άνθρωπος καλής πίστεως = καλόπιστος άνθρωπος
(personne de bonne foi)

έργο κοινής ωφελείας = κοινωφελές έργο
(travail d'utilité publique)

- iii) Réduction du modifieur adjectival, notée $N1$ ($E+DET:G$) ($E + A:G$) $N2:G$. Nous avons observé que dans un nombre assez important de $N1$ ($E+DET:G$) $A:G$ $N2:G$ du grec moderne, le modifieur adjectival est optionnel⁸⁴. La variante réduite constitue un nom composé binaire de structure N ($E+DET:G$) $N:G$. A titre d'illustration, citons les exemples suivants :

δελτίο (E + αστυνομικής) ταυτότητας

⁸⁴ La réduction du modifieur adjectival est un phénomène qui touche également d'autres types de noms composés complexes (ou surcomposés) du grec moderne. La forme réduite du nom composé complexe (ou surcomposé) est de structure N ($E+DET:G$) $N:G$. A titre indicatif, citons les exemples suivants :

$(E + A) N1$ ($E+DET:G$) $N2:G$ = N ($E+DET:G$) $N:G$

(E + απλή) μέθοδος των τριών

((E + simple) méthode de trois)

$N1$ ($E+DET:G$) $N2:G$ ($E + N3:G$) = N ($E+DET:G$) $N:G$

τόζο πορείας (E + οχήματος)

(clignotant)

L'étude de ce type de variantes réduites ne fait pas l'objet de la présente étude. Leur étude systématique et la description de leurs propriétés linguistiques dans les tables du lexique-grammaire reste donc à faire.

(carte d'identité)

ισοζύγιο (E + εξωτερικών) πληρωμών

(balance des paiements)

καθαγιασμός (E + των) (E + Τιμίων) Δώρων

(consécration des offrandes sacrées)

μάτι (E + της) (E + ηλεκτρικής) κουζίνας

(chauffe-plats)

ονομασία (E + ανωτέρας) προελεύσεως

(appellation d'origine contrôlée)

οικονομία (E+ της) (E + ελεύθερης) αγοράς

(économie du marché)

παιδί του (E + δοκιμαστικού) σωλήνα

(bébé-éprouvette)

- iv) Réduction de l'adjectif et du deuxième composant nominal, notée $N1 (E+DET:G) A:G N2:G = N1$. Comme nous l'avons noté pour certains noms composés binaires de structure $N (E+DET:G) N:G$, il est possible que certains noms composés complexes (ou surcomposés) présentent une variante de nom simple. Cependant, il s'agit d'un phénomène marginal dans le cadre des noms composés complexes (ou surcomposés). A signaler que la variante réduite présente des problèmes d'ambiguïté (cf. II, 5.2.5.3). A titre indicatif, citons l'exemple suivant :

δίκτυο (E + ηλεκτρονικών υπολογιστών)

(réseau (E + d'ordinateurs))

- v) Réduction du nom-tête et réduction optionnelle du deuxième composant nominal, notée $N1 (E+DET:G) A:G N2:G = A:G (E + N2:G)$. Dans ce cas, les variantes réduites constituent respectivement un nom composé de structure AN ou mot simple (adjectif substantivé). Les variantes réduites restent invariables au génitif (cf. II, 5.2.5.2). Nous n'avons rencontré qu'un seul exemple de ce type :

δελτίο ελευθέρας εισόδου = ελευθέρας (E + εισόδου)

(bulletin d'accès libre)

- vi) L'adjectif du deuxième complément nominal est antéposé, variante notée *NI* (*E+DET:G*) *A:G* *N2:G* = *A* *NI* (*E+DET:G*) *N2:G*. Dans ce cas, la variante syntaxique constitue un nom composé complexe (ou surcomposé). Ce phénomène n'est pas très fréquent. A titre indicatif, citons les exemples suivants :

NI *A:G* *N2:G* *A* *NI* *N2:G*

μέσο μαζικής μεταφοράς = *μαζικό μέσο μεταφοράς*

(moyen de transport)

μέσα μαζικής ενημέρωσης = *μαζικά μέσα ενημέρωσης*

(communications de masse)

- vii) Le *A:G* du *A:G* *N2:G* accepte une paraphrase en *N2:G*, variante notée *NI* (*E+DET:G*) *N:G-a* *N2:G* = *NI* (*E+DET:G*) *N2:G* (*E+DET:G*) *N3:G*. Il s'agit de l'adjectivation du deuxième complément nominal au génitif et de la co-occurrence des deux formes variées. Nous avons étudié ce type de variante dans le cadre des noms composés binaires (cf. II, 5.2.5.5). Le *A* est un adjectif morphologiquement associé au *N3:G*.

NI *N:G-a* *N2:G*

διακήρυξη των ανθρωπίνων δικαιωμάτων

NI *N2:G* *N3:G*

= *διακήρυξη των δικαιωμάτων του ανθρώπου*

(déclaration des droits de l'homme)

Des points de vue syntaxique et morphologique, l'adjectivation du *N3:G* est postérieure à la formation du groupe nominal en *N:G*. Pour cette raison, nous considérons que cette propriété doit être représentée de manière formalisée dans la table du lexique-grammaire consacrée aux *NI* (*E+DET:G*) *N2:G* (*E+DET:G*) *N3:G*.

- viii) Antéposition de l'adjectif du *A:G* *N2:G* et effacement du *N2:G*, notée *NI* (*E+DET:G*) *A:G* *N2:G* = *A* *NI*. La variante réduite constitue un nom composé binaire de type *AN* avec le même nom-tête. Ce phénomène est peu courant dans les

noms composés complexes (ou surcomposés) de structure $N (E+DET:G) N:G = NI (E+DET:G) A:G N2:G$. A titre d'illustration, citons les exemples suivants :

$NI \quad A:G \quad N2:G \quad A \quad NI$

λειτουργία των προηγιασμένων δώρων = προηγιασμένη λειτουργία
(messe des Présanctifiés)

$NI \quad A:G \quad N2:G \quad A \quad NI$

δέσμη φωτεινών ακτίνων = φωτεινή δέσμη
(faisceau lumineux)

$N \quad A:G \quad N2:G \quad A \quad NI$

επιχείρηση μικτής οικονομίας = μικτή επιχείρηση
(entreprise d'économie mixte)

- ix) Réduction du nom-tête qui donne lieu à la formation d'un nouveau AN , notée $NI (E+DET:G) A:G N2:G = A N2$. Le AN formé est identique au $A:G N2:G$ du nom composé complexe (ou surcomposé) de structure $N (E+DET:G) N:G = NI (E+DET:G) A:G N2:G$, mais il ne reste pas invariable au génitif : il se décline en cas et/ou en nombre. Ce type de variante n'est pas très courant. A titre indicatif, citons les exemples ci-dessous :

$NI \quad A:G \quad N2:G \quad A \quad N2$

αγώνας ανώμαλου δρόμου = ανώματος δρόμος
(course tout-terrain)

$NI \quad A:G \quad N2:G \quad A \quad N2$

δελτίο αστυνομικής ταυτότητας = αστυνομική ταυτότητα
(carte d'identité)

- x) La permutation, notée $NI (E+DET:G) A:G N2:G = A:G N2:G (E+DET) NI$. Nous n'avons rencontré qu'un cas unique de permutation dans les noms composés complexes (ou surcomposés) de structure $NI (E+DET:G) A:G N2:G$:

$NI \quad A:G \quad N2:G \quad A:G \quad N2:G \quad NI$

κάθειρξη αόριστης διάρκειας = αόριστης διάρκειας κάθειρξη
(emprisonnement à durée indéterminée)

Nous avons effectué la représentation formalisée des variantes syntaxiques des *N1 (E+DET:G) A:G N2:G* au moyen des tables du lexique-grammaire (cf. V, 11.3.3)

En guise de conclusion, signalons que les différents types de variantes affectent 36% des noms composés complexes (ou surcomposés) de structure *N1 (E+DET:G) A:G N2:G*.

TROISIEME PARTIE : STRUCTURE LEXICALE INTERNE

Chapitre 6. Les différentes catégories et sous-catégories des $N(E + DET:G)N:G$

6.1 Introduction

Après avoir complété le recensement des données lexicales (cf. II, 4.2.1-4.2.3) et l'étude des différents phénomènes de variation (cf. III, 5), nous nous intéressons à étudier leur structure lexicale interne. Cette démarche nous permettra de délimiter les unités lexicales qui entrent dans ces noms composés et qui mettent éventuellement en jeu des phénomènes lexicaux et grammaticaux particuliers. A noter que la description de la structure lexicale interne des $N(E + DET:G)N:G$ s'avère indispensable pour leur représentation formalisée dans le dictionnaire électronique morphologique, dans les grammaires locales ou dans les tables du lexique-grammaire (cf. V, 11.3.1-11.3.3).

Dans cette partie, nous cherchons donc à étudier la structure interne de l'ensemble des structures faisant l'objet de cette recherche, qu'elles soient des noms composés binaires ou des noms composés complexes (ou surcomposés). Rappelons ici que ces structures peuvent correspondre à des :

- noms composés entièrement figés,
- syntagmes nominaux semi-figés (ou productifs),
- entités nommées,
- termes de spécialité,
- syntagmes nominaux figés faisant partie d'une phrase simple figée et qui ont également un emploi autonome en dehors de l'expression figée.

De manière générale, les phénomènes observés et décrits dans cette partie affectent toutes les sous-catégories précitées. Des phénomènes spécifiques de chaque sous-catégorie (notamment, des entités nommées et des syntagmes nominaux semi-figés (ou productifs)) seront signalés. Plus particulièrement, nous examinons la nature morphosyntaxique des constituants des noms

composés. Mais, avant d'aborder la structure interne des noms composés, il nous paraît utile de déterminer et délimiter les différentes sous-catégories que nous avons repérées.

6.1 Les noms composés binaires de structure $N (E+DET:G) N:G$

Comme nous l'avons déjà vu, les noms composés de structure $N (E+DET:G) N:G$ peuvent être binaires ou complexes (ou surcomposés) (cf. II, 4.2.2). Les noms composés binaires sont constitués de deux composants nominaux « pleins », à savoir le nom-tête et le deuxième composant nominal au génitif et son déterminant (le déterminant peut être zéro). Le déterminant s'accorde obligatoirement en cas, genre et nombre avec le deuxième composant. Les deux composants nominaux peuvent être des substantifs ou des adjectifs substantivés, comme par exemple :

$N+A \quad DET:G \quad N:G=:$

εσχάτη των ποινών

(peine capitale)

$N \quad DET:G \quad N+A:G=:$

καλάθι των αχρήστων

(poubelle)

Du point de vue flexionnel, seul le nom-tête varie en cas et en nombre alors que le deuxième composant nominal constitue une constante lexicale au génitif (singulier ou pluriel)⁸⁵. Le genre du nom composé binaire est hérité du genre du nom-tête.

6.3 Les noms composés complexes (ou surcomposés)

Par noms composés complexes (ou surcomposés) nous entendons les expansions des noms composés binaires de structure $N (E+DET:G) N:G$, c'est-à-dire les noms composés dont au moins un de leurs constituants nominaux est lui aussi un nom composé binaire. Précisons que les noms composés binaires faisant partie des noms composés complexes (ou surcomposés) doivent figurer comme entrées autonomes dans le dictionnaire électronique morphologique des noms composés et dans les tables du lexique-grammaire correspondantes.

⁸⁵ Comme nous l'avons discuté dans II, 5.2.1.6, dans certains $N (E+DET:G) N:G$, la réalisation du N:G est possible au singulier ou au pluriel sans changement de sens.

Du point de vue flexionnel, les noms composés complexes (ou surcomposés) de structure $N (E+DET:G) N:G$ se comportent de la même manière que les noms composés binaires (cf. III, 6.2). Du point de vue de leur constitution lexicale interne, ils peuvent être définis par les structures morphosyntaxiques suivantes⁸⁶ :

i) $N1 (E+DET:G) A:G N2:G =:$

ψάρι του γλυκού νερού

poisson la-*Gns* douce-*Gns* eau-*Gns*

(poisson d'eau douce)

ώρες κοινής ησυχίας

heures commun-*Gfs* silence-*Gfs*

(heures de repos)

ii) $N1 (E+DET:G) N2:G (E+DET:G) N3:G =:$

άδεια άσκησης επαγγέλματος

permis exercice-*Gfs* profession-*Gns*

(autorisation d'exercice)

διακήρυξη των δικαιωμάτων του ανθρώπου

déclaration les-*Gnp* droits-*Gnp* le-*Gms* homme-*Gms*

(déclaration des droits de l'homme)

ωράριο λειτουργίας (Ε + των) καταστημάτων

horaire ouverture-*Gfs* (E + les-*Gnp*) magasins-*Gnp*

(horaire d'ouverture des magasins)

iii) $A N1 (E+DET:G) N2:G =:$

μέσος όρος αριθμών

moyen terme chiffres-*Gmp*

(moyenne)

μεγάλη κυκλοφορία του αίματος

⁸⁶ Nous ne citons ici que des exemples de noms composés complexes (ou surcomposés) qui comportent jusqu'à quatre mots simples « pleins ».

grande circulation le-*Gns* sang-*Gns*
(grande circulation sanguine)

Notons que les noms composés complexes (ou surcomposés) ainsi décomposés peuvent être classés soit comme $N (E+DET:G) N:G$ soit comme $A N$. Il n'est pas toujours facile de décider où les classer à cause du chevauchement de noms composés binaires à l'intérieur du nom composé complexe (ou surcomposé). Par exemple, le nom composé complexe (ou surcomposé) :

ανώτατο όριο ταχύτητας
maximale limite vitesse-*Gfs*
(limite de vitesse)

donne lieu à une décomposition qui peut être effectuée des deux façons suivantes :

$AN =:$	N	$N:G$
ανώτατο όριο		ταχύτητας
maximale limite		vitesse- <i>Gfs</i>

ou bien

A	$N =:$	$N N:G$
ανώτατο		όριο ταχύτητας
maximale		limite vitesse- <i>Gfs</i>

dans la mesure où nous acceptons que *ανώτατο όριο*/limite maximale et *όριο ταχύτητας*/limite de vitesse constituent tous les deux des noms composés binaires de structure AN et $N (E+DET:G) N:G$ respectivement.

iv) $N1 (E+DET:G) N2:G CONJC (E+DET:G) N3:G =:$

ανάπαυση ψυχής και σώματος
repos âme-*Gfs* et corps-*Gns*
(repos en paix)

δικαίωμα του εκλέγειν και του εκλέγεσθαι
droit le-*Gns* élire-*Gns* et le-*Gns* [être élu]-*Gns*
(droit d'élire et d'être élu)

v) $A\ N1\ (E+DET:G)\ N2:G\ (E+DET:G)\ N3:G =:$

γενικός δείκτης τιμών καταναλωτή

général indice prix-_{Gfp} consommateur-_{Gms}

(indice des prix à la consommation)

vi) $A\ N1\ (E+DET:G)\ A:G\ N2:G =:$

νομικό πρόσωπο δημοσίου δικαίου

juridique personne public-_{Gns} droit-_{Gns}

(personne morale de droit public)

vii) $N1\ (E+DET:G)\ A1:G\ A2:G\ N2:G =:$

συνδετήρας εναέριων ηλεκτρικών αγωγών

lien aériens-_{Gmp} électriques-_{Gmp} tubes-_{Gmp}

(connecteur de lignes électriques aériennes)

viii) $N1\ (E+DET:G)\ ADV\ A:G\ N2:G =:$

ρήτρα του μάλλον ευνοούμενου κράτους

clause le-_{Gns} plus favorisé-_{Gns} Etat-_{Gns}

(clause de la nation la plus favorisée)

ix) $N1\ (E+DET:G)\ N2:G\ (E+DET:G)\ N3:G\ CONJC\ (E+DET:G)\ N4:G =:$

δέντρο της γνώσης του καλού και του κακού

arbre la-_{Gfs} connaissance-_{Gfs} le-_{Gns} bien-_{Gns} et le-_{Gns} mal-_{Gns}

(arbre de vie)

Parmi les différentes structures morphosyntaxiques présentées ci-dessus, nous avons décidé d'examiner et décrire systématiquement dans le cadre de la présente étude (cf. II, 5.3) les $N1\ (E+DET:G)\ A:G\ N2:G$ car ils étaient les plus nombreux dans notre corpus : nous avons identifié 244 $N1\ (E+DET:G)\ A:G\ N2:G$ dans un ensemble de 484 noms composés complexes (ou surcomposés), ce qui représente plus de 50% des noms composés complexes recensés.

Notons enfin que les noms composés binaires et les noms composés complexes (ou surcomposés) sont répartis à leur tour en cinq sous-catégories dont les principales caractéristiques sont présentées dans les sections suivantes (cf. III, 6.4-6.8).

6.4 Les noms composés entièrement figés

Comme nous l'avons vu dans I, 2.3.1, les noms composés entièrement figés (binaires ou surcomposés) sont ceux qui ne permettent aucune variation d'ordre, notamment, syntaxique et lexical. A titre indicatif, citons un exemple pour lequel nous observons que toute sorte de variation est interdite :

- au niveau lexical

- la substitution du complément nominal par d'autres substantifs sémantiquement voisins n'est pas possible :

μερίδα του λέοντος

(part du lion)

**μερίδα (της τίγρης + του ελέφαντα + του φιδιού +...)*

*part (du tigre + de l'éléphant + du serpent +...)

- la substitution de la forme savante par la forme démotique est impossible :

μερίδα του λέοντος

(variante savante)

**μερίδα του λιονταριού*

(variante démotique)

- au niveau syntaxique

- les variantes de structure et transformationnelles ne sont pas acceptables :

μερίδα του λέοντος

(N DET:G N:G)

part le lion-*Gms*

(part du lion)

**μερίδα λέοντος*

(réduction du déterminant, DET:G = E)

*part lion-*Gms*

<i>*λεόντεια μερίδα</i>	(adjectivation du complément de nom,
<i>*léonine part</i>	N DET:G N:G = N:G-a N)
<i>*λεοντομερίδα</i>	(soudure, N DET:G N:G = Ns)
<i>*lionpart</i>	

- au niveau phonologique

άνθος του αγρού	
<i>*άνθος τ' αγρού</i>	(élision)
(fleur des champs)	

Mais les noms composés entièrement figés sont rares. Nous avons repéré des structures qui donnent lieu à plusieurs types de variation et d'autres qui donnent lieu à un paradigme lexical productif. Nous examinons ces structures dans la section suivante.

6.5 Les syntagmes nominaux semi-figés (ou productifs)

Rappelons que dans le cadre de la présente étude, sont considérés comme syntagmes nominaux semi-figés (ou productifs) les constructions dont :

- i) le nom-tête peut être combiné avec un grand nombre de noms au génitif :

αγορά (αξιών + εργασίας + κεφαλαίου + μετοχών + συναλλάγματος + χρήματος + αμοιβαίων κεφαλαίων + ...)
 (marché (de valeurs + du travail + des capitaux + des actions + des changes + monétaire + des fonds + ...))

- ii) le composant nominal au génitif s'applique à un large spectre des noms-têtes :

(βιβλίο + λεξικό + υπολογιστής + ...) τσέπης
 ((livre + dictionnaire + calculatrice + ...) de poche)

Les syntagmes nominaux semi-figés (ou productifs) peuvent être binaires :

άδεια (οπλοφορίας + παραμονής + κατεδάφισης + υλοτομίας + υπέρπτησης + κυνηγιού + ταξί + ...)
 (permis (de port d'armes + de séjour + de démolition + pour l'abattage d'arbres + de survol + de chasse + de taxi + ...))

ou bien complexes (ou surcomposés), comme en témoigne l'exemple ci-dessous :

άδεια (άσκησης επαγγέλματος + κινητής τηλεφωνίας + ελεύθερης κυκλοφορίας + παροχής υπηρεσιών + ...)

(permis (d'exercice de profession + de téléphonie mobile + de libre circulation + de prestation de services))

Rappelons⁸⁷ ici que nous pouvons subdiviser les syntagmes nominaux semi-figés (ou productifs) en deux catégories en fonction de leur variabilité :

- 1) les syntagmes nominaux semi-figés (ou productifs) permettant un nombre de variantes restreint et, éventuellement, calculable.

Concernant cette catégorie, les variantes observées se situent au niveau lexical ou orthographique⁸⁸ et peuvent affecter le nom-tête ou le deuxième composant nominal au génitif. A titre indicatif, citons quelques exemples susceptibles de variations de :

nom-tête :

(εταιρεία + εταιρία) χαρτοφυλακίου

(société à portefeuille)

deuxième complément nominal :

αγορά (κεφαλαίου + κεφαλαίων)

(marché des capitaux)

Les variantes lexicales et orthographiques sont numériquement restreintes et sémantiquement équivalentes ; il est donc aisé de les recenser et de les représenter dans le dictionnaire électronique morphologique et dans les tables du lexique grammair.

- 2) les syntagmes nominaux semi-figés (ou productifs) à complément de nom libre.

⁸⁷ Pour les syntagmes nominaux semi-figés (ou productifs), cf. aussi I, 2.3.2.

⁸⁸ Pour une étude plus détaillée des variantes lexicales et orthographiques, cf. II, 5.2.1 et 5.2.2.

Enfin, les syntagmes nominaux semi-figés (ou productifs) de cette catégorie se caractérisent par la présence d'un deuxième complément nominal lexicalement productif (cf. I, 2.3.2). A titre indicatif, citons l'exemple suivant :

γραφείο (αλλοδαπών + πληροφοριών + δευτεροβάθμιας εκπαίδευσης + δημοσίων σχέσεων + διανομής + εισαγωγής + εργασίας + μελετών + παραγωγής + προώθησης + στοιχημάτων + τύπου + απολεσθέντων + περιβάλλοντος + σταδιοδρομίας + ...)

(bureau (d'étrangers + de renseignements + d'enseignement secondaire + de relations publiques + de distribution + d'importation + de travail + d'études + de production + de promotion + de paris + de presse + d'objets perdus + d'environnement + de carrière + ...))

De manière générale, les syntagmes nominaux semi-figés (ou productifs) se caractérisent par l'absence de déterminant (déterminant zéro).

La méthode de représentation qui a été choisie pour les syntagmes nominaux semi-figés (ou productifs) dont le paradigme est ouvert sont les graphes lexicalisés construits manuellement (cf. V, 11.3.2). La localisation des syntagmes nominaux semi-figés (ou productifs) dans les textes peut se faire par l'application de ceux-ci.

6.6 Les entités nommées

L'étude systématique des entités nommées n'était pas en soi l'objectif de notre recherche, mais nous avons dû nous confronter à ce sujet étant donné qu'un nombre important des noms composés binaires ou complexes (ou surcomposés) de structure $N(E+DET:G)N:G$ correspondent à des entités nommées d'organisations, de lieux, d'événements, d'œuvres etc. (cf. II, 4.2.1). Ainsi, dans le cadre de la présente étude, nous nous sommes tout simplement contentée d'appliquer un modèle typologique, afin de distinguer les entités nommées des autres sous-catégories des $N(E+DET:G)N:G$ examinées ici.

« On désigne traditionnellement par le terme d'entité nommée les noms propres au sens classique, les noms propres dans un sens élargi mais aussi les expressions de temps et de quantité » (C. Martineau *et al.* 2007). Par conséquent, « les entités nommées désignent les noms de personnes, de lieux, d'organisations mais aussi les dates ou les unités monétaires. Alors que ces éléments ont longtemps été délaissés par les systèmes de traitement automatique des langues, le renouveau lié au travail sur corpus a révélé qu'il s'agissait en fait

d'éléments majeurs pour l'analyse. Les conférences en extraction d'information (*Message Understanding Conferences*, cf. MUC-7, 1998) ont mis en avant plusieurs tâches génériques, au premier rang desquelles l'analyse des entités nommées » (T. Poibeau 2005).

Comme le note C. Martineau *et al.* 2007, « plusieurs typologies ont été proposées pour les noms propres en général ; certaines morphosyntaxiques (D. Allerton 1987), d'autres graphiques (B. Daille et E. Morin 2000) ou bien encore sémantiques (T. Grass 2000). Pour ce qui est des entités nommées, d'autres catégorisations ont été établies, parmi lesquelles nous citons ESTER (C. Le Meur, S. Galliano et E. Geoffrois 2004) et celle de S. Sekine *et al.* 2002 ». Dans le cadre de notre étude, et dans le but d'effectuer une première classification des entités nommées, nous adoptons la typologie générale de C. Martineau *et al.* 2007. Elle comporte les neuf classes d'entités nommées suivantes :

Noms d'organisations :

Υπουργείο Δικαιοσύνης
(Ministère de Justice)

Εθνική Τράπεζα της Ελλάδος
(Banque Nationale de Grèce)

Noms de lieux (ou toponymes) :

Γη του Πυρός
(Terre de Feu)

Ηνωμένες Πολιτείες Αμερικής
(Etats-Unis d'Amérique)

Noms de faits :

μάχη του Μαραθώνα
(bataille de Marathon)

Ολυμπιακοί Αγώνες του 2004
(Jeux Olympiques de 2004)

Noms d'œuvres :

Ο τελευταίος των Μοϊκανών
(le dernier des Mohicans)

Το πορτρέτο του Ντόριαν Γκρέι
(le portrait de Dorian Gray)

Dates et heures (ou expressions temporelles) :

Κυριακή του Πάσχα
(dimanche de Pâques)

Εβδομάδα των Αγίων Παθών
(semaine sainte)

Noms de personnes (ou anthroponymes) :

Γέρος του Μοριά
(Vieux de Morée)

Κυρά της Ρω
(Dame de Ro)

Expressions numériques :

θερμοκρασία τήξεως (Ε + του πάγου)
(température de fusion (E + de la glace))

θερμοκρασία βρασμού (Ε + του ύδατος)
(température d'ébullition (E + de l'eau))

Coordonnées :

Nous n'en avons pas rencontré dans notre corpus des *N (E+DET:G) N:G* (cf. II, 4.2.1-4.2.3)

Il existe actuellement de nombreux systèmes efficaces de repérage, d'extraction et d'analyse d'entités nommées en différentes langues. Certains reposent sur une approche symbolique et utilisent une base de règles et de ressources lexicales construites à la main (J. Hobbs *et al.* 1996). D'autres sont fondés sur une approche statistique et emploient des connaissances

acquises automatiquement par apprentissage (D. M. Bikel *et al.* 1997 ; S. Sekine et C. Nobata 1998). D'autres, enfin, adoptent une approche hybride (P. Watrin 2006).

Dans le cas d'extracteurs d'entités nommées utilisant une approche symbolique, l'identification des entités nommées peut être considérablement améliorée par l'exploitation du contexte immédiat ou éloigné. Dans le cadre de cette étude, nous avons constaté que les séquences qualifiées de syntagmes nominaux semi-fixés (ou productifs) constituent dans plusieurs cas le contexte immédiat gauche d'une entité nommée. Ils s'avèrent donc très utiles pour le repérage et l'identification des entités nommées dans les corpus. Ils peuvent alors constituer des « preuves internes » (D. McDonald 1996 : 22). Dans ce cas, ils font partie de l'entité nommée et sont capitalisés, comme par exemple :

Κέντρο Ερευνών (Προοδευτικής Πολιτικής + Παράδοσης και Πολιτισμού + ...)
Centre de recherche (de politique progressive + de tradition et de culture + ...)

Ils peuvent également constituer des « preuves externes » (D. McDonald 1996 : 23). Dans ce cas, ils n'appartiennent pas forcément à l'entité nommée : ils constituent le contexte d'apparition qui déclenche sa reconnaissance et ne sont pas capitalisés. Par exemple :

κέντρο ερευνών (Intercollege + Ρίζες + ...)
centre de recherche (Intercollege + Racines + ...)

Dans le cadre de cette étude, nous avons adopté une solution provisoire : nous avons constitué une liste séparée des entités nommées de structure $N (E+DET:G) N:G$ recensées, tout en indiquant leur type (cf. Annexe 6). Une telle liste pourrait être utilisée dans des systèmes de repérage, d'extraction ou d'analyse d'entités nommées qui reposent sur l'approche linguistique. Dans le cas où l'on opte pour l'introduction des entités nommées dans le dictionnaire électronique morphologique des $N (E+DET:G) N:G$, nous proposons d'assigner à ces entrées un code spécifique, comme ceci a été fait pour les termes de spécialité (cf. III, 6.7 et V, 11.3.1.1). Par exemple : **EN-org** pour indiquer qu'il s'agit d'une entité nommée de type « organisation ». Ceci faciliterait l'extraction de cette partie du lexique et son exploitation dans des applications spécifiques. Nous envisageons d'effectuer ce codage dans une étude ultérieure.

6.7 Les termes de spécialité

« Une langue donnée est constituée par un ensemble diversifié de sous-codes que le locuteur emploie en fonction de ses modalités dialectales et qu'il sélectionne en fonction de ses besoins d'expression et selon les caractéristiques de chaque situation de communication. Cependant, au-delà de cette diversité foisonnante, toute langue possède un ensemble d'unités et de règles que tous ses locuteurs connaissent. Cet ensemble de règles, d'unités et de restrictions qui font partie des connaissances de la majorité des locuteurs d'une langue constitue ce qu'on appelle la langue commune ou générale » (T. Cabré 1998 : 115).

En effet les langues de spécialité ne sont pas à dissocier de la langue générale car elles en constituent un sous-ensemble. Les termes de spécialité constituent donc le continuum du lexique de la langue générale. Les langues de spécialité utilisent le même système de règles syntaxiques et grammaticales que la langue générale. C'est au niveau sémantique que se note la distinction entre la langue générale et une langue de spécialité. « Une langue spécialisée ne se réduit pas à une terminologie : elle utilise des dénominations spécialisées (les termes), y compris des symboles non linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée. On peut donc la définir comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées (P. Lerat 1995 : 21).

Ce travail ne porte pas sur l'étude approfondie des termes de spécialité et, pour cette raison, nous ne prenons pas de décision sur l'éventuel statut terminologique des $N(E+DET:G)N:G$ recensés. Cependant, il nous a paru utile de fournir l'information sur le domaine de spécialité lorsque celle-ci est enregistrée dans les dictionnaires usuels que nous avons compilés (cf. II, 4.2.2). A titre indicatif, citons quelques exemples des termes de spécialité de structure $N(E+DET:G)N:G$:

σημείο ανάφλεξης

(point d'inflammation)

(domaine : Chimie)

μήκος κύματος

(longueur d'onde)

(domaine : Physique)

μοχλός ταχυτήτων

(levier de vitesse)

(domaine : Technologie)

<i>τηλεόραση υψηλής ευκρίνειας</i> (télévision à haute définition)	(domaine : Electronique)
<i>αποκόλληση του αμφιβληστροειδούς χιτώνα</i> (décollement de la rétine)	(domaine : Médecine)
<i>διατάραξη της δημόσιας τάξης</i> (perturbation de l'ordre public)	(domaine : Droit)

Nous proposons l'ajout du trait *TS* pour les entrées du dictionnaire électronique morphologique qui constituent des termes de spécialité (cf. V, 11.3.1.1). A l'heure actuelle, nous avons effectué ce codage uniquement dans les tables du lexique-grammaire (cf. V, 11.3.3). Cependant, nous envisageons d'ajouter cette information dans le dictionnaire électronique morphologique lors d'une étude ultérieure.

6.8 Les syntagmes nominaux figés faisant partie d'une phrase simple figée et qui ont également un emploi autonome en dehors de celle-ci

L'étude des phrases simples figées ne constituant pas l'objet de notre travail, nous n'en effectuons pas une étude approfondie et détaillée ici⁸⁹. Dans le cadre de cette étude, nous nous intéressons uniquement aux $N(E+DET:G)N:G$ qui font partie d'une phrase simple figée, mais qui peuvent également avoir un emploi autonome en dehors de celle-ci. Par exemple, le $N(E+DET:G)N:G = :$ *βάπτισμα του πυρός*/baptême du feu peut apparaître aussi de manière autonome dans des phrases compositionnelles comme :

Το βάπτισμα του πυρός αργεί

Le baptême du feu tarde

Το βάπτισμα του πυρός ήρθε με την Αγία Πεντηκοστή

Le baptême du feu est arrivé avec la sainte pentecôte

Ces $N(E+DET:G)N:G$ constituent des entrées de notre dictionnaire électronique morphologique.

⁸⁹ Pour une étude plus approfondie, nous renvoyons aux travaux de J. Giry-Schneider 1987 et A. Fotopoulou 1993.

En revanche, les $N(E+DET:G)N:G$ comme *στήλη άλατος*/stèle sel-*Gns*, qui n'ont pas d'emploi autonome en dehors de la phrase simple figée ne sont pas inclus dans nos données. Nous considérons que ces $N(E+DET:G)N:G$ devraient être étudiés dans le cadre des expressions figées.

Chapitre 7. La structure lexicale interne des noms composés $N (E+DET:G) N:G$

7.1 Introduction

Après avoir explicité les différentes sous-catégories des $N (E+DET:G) N:G$ du grec moderne, nous nous proposons d'étudier leur structure lexicale interne. De manière générale, les noms intervenant dans les $N (E+DET:G) N:G$ se comportent de la même façon que dans les groupes nominaux libres. En revanche, les déterminants présentent des particularités qui sont spécifiques au phénomène du figement. Nous en discutons en détail dans les sections suivantes.

7.2 Les déterminants

7.2.1 Les déterminants dans les noms composés de structure $N (E+DET:G) N:G$

Plusieurs études ont été consacrées aux déterminants en général parmi lesquelles nous citons celles des I. Tsamadou-Jacobberger (1989, 1993, 2002), Z. Gavrilidou (1998, 2001, 2002), M. Setatos (1992), F. Corblin (1983, 1987, 1989), J.-C. Anscombre (1991), G. Kleiber (1985, 1989, 1990, 1995, 1998, 2001), P.-A. Buvet (1993, 1994, 1998, 2001), X. Blanco (2002), G. Gross (1986a, 1995a, 2002), M. Gross (1985, 2001b, 2002) et D. Le Pesant (2002).

Nous focaliserons notre intérêt sur les déterminants dans les noms composés de structure $N (E+DET:G) N:G$ du grec moderne. X. Blanco (2001 : 70, 73) souligne que « les rapports entre détermination et figement affectent aussi bien le fonctionnement des déterminants à l'intérieur des structures figées que les déterminants figés eux-mêmes ». Il importe donc de distinguer, d'une part, les déterminants, qui sont figés par rapport au N qu'ils accompagnent et, d'autre part, les déterminants composés figés (ou complexes), qui le sont indépendamment du N déterminé.

Les noms composés commentés ici ont un déterminant plus ou moins figé. Nous nous intéressons ici au déterminant à l'intérieur du nom composé. Etant donné que les $N (E+DET:G) N:G$ sont définis par des contraintes syntaxiques, lexicales et sémantiques, leurs déterminants mis en jeu sont en principe sémantiquement vides et syntaxiquement non-

actualisés. Toutefois, les $N(E+DET:G)N:G$ grecs sont susceptibles de diverses variations de leurs composants déterminants.

Nous cherchons à étudier le type et la distribution des déterminants dans les combinaisons lexicales des $N(E+DET:G)N:G$ grecs. Nous examinons, d'une part, les formes à déterminants définis et indéfinis et, d'autre part, celles à déterminant zéro ($Det=:E$).

Notons que, du point de vue morphologique, les déterminants du grec moderne se divisent en déterminants simples et composés⁹⁰, tout comme les autres catégories grammaticales de la langue. Du point de vue flexionnel, les déterminants du grec moderne se déclinent en nombre (singulier et pluriel), en genre (masculin, féminin, neutre) et en cas (nominatif, génitif, accusatif). Pour ce qui est des noms composés grecs de structure $N(E+DET:G)N:G$, les déterminants s'accordent en cas, genre et nombre avec le deuxième composant nominal. Ils apparaissent donc uniquement au génitif, dans les trois genres et dans les deux nombres.

7.2.1.1 Déterminants définis

Pour les noms composés binaires de structure $N(E+DET:G)N:G$, nous avons observé les types de déterminants définis présentés ci-dessous :

- les articles définis ;
- les adjectifs possessifs sans source.

Nous présentons ci-après la nature morphologique de ces deux catégories de déterminants définis dans les noms composés de structure $N(E+DET:G)N:G$ du grec moderne.

7.2.1.1.1 Articles définis

L'article défini (noté *Ddéf*) est un des déterminants que nous avons le plus fréquemment rencontré au cours de cette étude. A titre indicatif, donnons ici quelques exemples où le déterminant du deuxième complément est un *Ddéf* :

κλειδί του μυστηρίου

clé le-*Gns* mystère-*Gns*

(clé du mystère)

στέφανος του μαρτυρίου

couronne le-*Gns* calvaire-*Gns*

⁹⁰ Cf. X. Blanco ; P.-A. Buvet et Z. Gavriilidou (1999).

(couronne d'épines)

*τσάι του βουνού*⁹¹

thé la-*Gms* montagne-*Gms*

(thé des montagnes)

*ψάρι του αφρού*⁹²

poisson le-*Gms* écume-*Gms*

(poisson d'écume)

ψήφος της Αθηνάς

vote la-*Gfs* Athéna-*Gfs*

(vote d'Athéna)

Notons que dans les noms composés l'article défini peut se combiner avec des infinitifs verbaux⁹³ ou des adjectifs substantivés⁹⁴, comme par exemple :

τρόπος του λέγειν

(mode de dire)

μανδύας των ψυχοπαθών

(camisole de force)

Dans nos données lexicales, l'article défini a l'emploi générique. Par exemple, dans :

τσάι του βουνού

(thé des montagnes)

⁹¹ Cette forme présente également une variante transformationnelle de type $N\ DET:G\ N:G = \langle N:G-a \rangle\ N$. Les variantes de ce type sont étudiées dans II, 5.2.5.5.

⁹² Cette forme présente également une variante transformationnelle de type $N\ DET:G\ N:G = Ns$. Les variantes de ce type sont étudiées dans II, 5.2.5.6.

⁹³ L'infinitif a disparu depuis longtemps du système actuel verbal du grec moderne (Dictionnaire de l'Institut des Études Néohelléniques 1998 : 229-230) et son utilisation est limitée aujourd'hui uniquement à des structures figées. Pour un certain nombre de verbes, l'infinitif peut également apparaître substantivé (*Ddéf Vinf*) dans des positions syntaxiques argumentales (sujet et complément d'objet).

⁹⁴ Précisons, cependant, que les formes $N\ (E+DET:G)\ N:G$ où l'on est en présence d'un adjectif substantivé en position de deuxième constituant sont des variantes réduites de noms composés complexes (ou surcomposées). Par exemple :

$N\ (E+DET:G)\ A:G\ N:G = N\ (E+DET:G)\ A-n:G$
αποκόλληση (Ε + του) αμφιβληστροειδούς (Ε + χιτώνα)
(décollement de la rétine)

Les variantes syntaxiques des noms composés complexes (ou surcomposées) sont étudiées dans II, 5.3.4.

il est utilisé pour désigner le thé qui pousse dans les montagnes en général, et non pas sur une montagne particulière.

On retrouve également l'article défini devant le deuxième composant nominal, quand celui-ci est un nom de personne, comme par exemple :

μίτος της Αριάδνης
(fil d'Ariane)

αίτημα του Ευκλείδη
(axiome d'Euclide)

μήλο του Αδάμ
(pomme d'Adam)

ou un nom de lieu, comme par exemple :

τείχος του Βερολίνου
(mur de Berlin)

πύργος της Βαβέλ
(tour de Babel)

Quant à la nature sémantique du deuxième composant nominal, qui est déterminé par l'article défini, il peut être concret :

μάτι της βελόνας
(trou de l'aiguille)

ou, dans la majorité des cas, abstrait :

δέντρο της ζωής
(arbre de vie)

Enfin, notons que les formes avec le déterminant défini présentent souvent une variante réduite à déterminant zéro (cf. III, 7.2.5.1).

7.2.1.1.1 Possessifs sans source

Le possessif sans source est peu courant. On le retrouve surtout dans le cas des noms composés qui font partie d'une phrase simple figée (cf. A. Fotopoulou 1993a : 100-101) :

Η Μαρία πλήρωσε τα μαλλιά της κεφαλής της γι' αυτόν τον πίνακα

La Maria a payé les cheveux la-*Gfs* tête-*Gfs* elle-*Gfs* pour ce tableau

(Maria a payé ce tableau très cher)

Cependant, nous le retrouvons également dans des noms composés qui ont un emploi autonome. Par exemple :

άντρας της ζωής μου

(homme de ma vie)

γυναίκα των ονείρων μου

(femme de mes rêves)

κλέφτης της καρδιάς μου

(voleur de mon cœur)

Notons que l'adjectif possessif des noms composés précités est porteur de coréférence sur un constituant de la phrase où il apparaît. Le référent (ou antécédent ou portée) des adjectifs possessifs est représenté par un exposant numérique, noté $Poss^i$, qui renvoie à une position syntaxique (ou argument du prédicat) de la phrase, notée N_i . Par exemple, dans la phrase :

Η Μαρία συνάντησε τον άντρα της ζωής της

(Maria a rencontré l'homme de sa vie)

la représentation de coréférence (ou portée) sera notée $Poss^0$ (= : της/sa), qui se réfère à N_0 (= : Μαρία/Maria).

Du point de vue morphologique, les adjectifs possessifs du grec moderne se subdivisent en adjectifs possessifs simples et composés, suivant la distinction générale des mots simples et composés de M. Silberztein (1990). Les adjectifs possessifs simples (notés $Poss_s$) sont en effet les formes réduites (ou faibles) des pronoms personnels au génitif (*μου*/mon ou ma, *σου*/ton ou ta, *του*/son ou sa, *μας*/notre, *σας*/votre, etc.). Ils peuvent être antéposés ou postposés par rapport au nom déterminé. Dans le cadre de la présente étude, nous n'avons rencontré que des adjectifs possessifs postposés par rapport au nom déterminé.

7.2.1.2 Déterminants indéfinis

Dans le cadre de cette étude, nous avons rencontré les déterminants indéfinis qui regroupent les deux cas présentés ci-dessous :

- les articles et adjectifs indéfinis ;
- les déterminants numéraux.

Nous présentons ci-après ces deux catégories de déterminants définis dans les noms composés de structure $N (E+DET:G) N:G$ du grec moderne.

7.2.1.2.1 Articles et adjectifs indéfinis

Le nombre des noms composés de structure $N (E+DET:G) N:G$ comportant des articles indéfinis est très limité, ce qui n'est pas le cas en français. Cette différence s'explique par le fait qu'en grec moderne le déterminant zéro ($Dét=: E$) est chargé de certaines fonctions qui, en français, sont attribuées aux déterminants indéfinis⁹⁵. En guise d'illustration, citons l'exemple suivant :

όροι μιας κρίσης

(termes d'un jugement)

Dans l'exemple précité, le déterminant (*μιας/une*) du nom composé (*όροι μιας κρίσης*/termes d'un jugement) est un article indéfini. « En grec moderne, tout comme en français, l'article indéfini ($Dind=: ένας-μια-ένα/un(e)$) est ambigu avec le déterminant numéral ($Dnum=: ένας-μία-ένα/un(e)$). Cette ambiguïté, qui prête à confusion au premier abord, peut être levée à l'aide du mécanisme de la commutation (cf. I. Tsamadou-Jacobberger 1993) ».

Ainsi, dans l'exemple précité, le déterminant *ενός/un* ne peut pas être remplacé par d'autres déterminants numéraux (cf. III, 7.2.1.2.2) :

**όροι δύο κρίσεων*

(*termes de deux jugements)

Examinons maintenant l'exemple suivant :

ξενοδοχείο ενός αστέρος

(hôtel une étoile)

Dans ce cas, la substitution d'autres déterminants numéraux au déterminant *ενός*/un est possible :

ξενοδοχείο (δύο + τριών + τεσσάρων + πέντε) αστερών
(hôtel (deux + trois + quatre + cinq) étoiles)

7.2.1.2.2 *Déterminants numéraux*

Les déterminants numéraux du grec moderne (notés en général *Dnum*) se distinguent, du point de vue morphologique, en déterminants numéraux simples et déterminants numéraux composés (M. Silberztein 1990), tout comme les autres catégories grammaticales de la langue.

« Du point de vue graphique, ils présentent des formes alphabétiques, numériques (en chiffres arabes, latins et grecs anciens) et alphanumériques. De plus, ils se subdivisent en six catégories, en fonction de leur statut syntaxico-sémantique dans les combinaisons *Dnum_N* (libres ou figées), où ils apparaissent (cf. M. Triantaphyllidis 2000 : 274-285, D. Holton *et al.* 2000 : 106-110) » (S. Voyatzi 2006 : 105).

Nous allons par la suite étudier quelques exemples d'emplois de déterminants numéraux dans les noms composés grecs de structure *N (E+DET:G) N:G* dans le but de mettre en lumière des phénomènes linguistiques particuliers. Nous avons cependant observé que les déterminants numéraux se comportent en général de la même manière dans les combinaisons figées (*Dnum_C*) que dans les combinaisons libres (*Dnum_N*).

Les noms composés de structure *N (E+DET:G) N:G* comportant des déterminants numéraux sont peu nombreux. Les déterminants numéraux qui interviennent sont des cardinaux et des ordinaux. A titre indicatif, citons les exemples suivants :

i) déterminants numéraux cardinaux

ξενοδοχείο δύο αστερών
(hôtel deux étoiles)

⁹⁵ Notons, d'ailleurs, que les noms composés grecs de structure *N (E+DET:G) N:G* à déterminant indéfini présentent souvent une variante réduite à déterminant zéro, comme par exemple : *πυκνότητα* (*E + ενός*) *σώματος*/densité d'un corps. Ce type de variation est étudié dans II, 5.2.5.1.

εβδομάδα πέντε ημερών

(semaine de cinq jours)

εβδομάδα τριάντα πέντε ωρών

(semaine de trente-cinq heures)

ii) déterminants numéraux ordinaux

είδος πρώτης ανάγκης

(article de première nécessité)

πέμπτος τροχός της αμάξης

(cinquième roue du carrosse)

Dans les exemples précités, les déterminants numéraux intervenant peuvent apparaître en forme alphabétique, numérique ou alphanumérique.

Concernant leur représentation dans le dictionnaire électronique morphologique des noms composés et dans les tables du lexique-grammaire, nous avons provisoirement procédé à la multiplication des entrées (à savoir : forme à *Dnum* alphabétique, forme à *Dnum* numérique et forme à *Dnum* alphanumérique). Nous avons opté pour cette solution parce que le nombre des noms composés comportant un déterminant numéral est très restreint. Cependant, nous envisageons de relier les formes alphabétiques aux formes (alpha)numériques correspondantes dans le dictionnaire morphologique électronique afin d'éviter les entrées redondantes.

7.2.1.2.3 *Déterminant zéro*

Un nombre important de noms composés grec de structure $N(E+DET:G)N:G$ se caractérisent par l'absence de déterminant (déterminant zéro, noté $Dét=: E$)⁹⁶. En guise d'illustration, citons les exemples suivants :

ζώνη ασφαλείας

(ceinture de sécurité)

απόσταση ασφαλείας

⁹⁶ Notons, cependant, que souvent les formes sans déterminant sont des variantes réduites de formes avec déterminant. Les variantes de ce type sont étudiées dans II, 5.2.5.1.

(distance de sécurité)

εργολάβος κηδειών

(entrepreneur de pompes funèbres)

ιστορία αγάπης

(histoire d'amour)

ταξίδι αναψυχής

(voyage d'agrément)

Dans la plupart des cas, le déterminant zéro à l'intérieur du nom composé est indépendant de la détermination du nom-tête. I. Tsamadou (1984 : 188) remarque que « si le génitif, complément de nom, n'est pas déterminé, il y a de fortes chances qu'il s'agisse d'une seule unité lexicale ». Cependant, comme nous l'avons déjà montré, cette remarque ne se limite pas au déterminant zéro, étant donné qu'il existe des noms composés de structure $N (E+DET:G) N:G$ dans lesquels le déterminant du complément de nom au génitif est l'article défini ou indéfini (cf. III, 7.2.1.1.1 et 7.2.1.2.1).

De même, D. Holton *et al.* (2000 : 280) notent que le déterminant est zéro dans des constructions figées dont le deuxième complément « est abstrait et au génitif (e.g. *λόγος υπάρξεως*/raison d'être, *άδεια οδήγησεως*/permis de conduire, *δελτίο ταυτότητος*/carte d'identité) », tout en précisant que le complément au génitif comporte souvent la marque flexionnelle savante. Signalons que, d'après nos données lexicales, les exemples cités par D. Holton *et al.* (2000 : 280) acceptent la variante démotique aussi : *λόγος ύπαρξης*/raison d'être, *άδεια οδήγησης*/permis de conduire, *δελτίο ταυτότητας*/carte d'identité. De plus, les entrées de notre dictionnaire ne nous permettent pas de dégager des conclusions sur la nature sémantique des composants nominaux. Dans les données lexicales que nous avons rassemblées, nous observons que le déterminant peut être zéro quand le premier substantif est concret et le deuxième composant nominal est abstrait. A titre indicatif, citons les exemples suivants :

ταινία ασφαλείας

(fil de sécurité)

σοκολάτα υγείας

(chocolat noir)

Mais le déterminant peut également être zéro quand les deux composants nominaux sont abstraits, par exemple :

συνωμοσία σιωπής
(complot du silence)

ποιότητα ζωής
(qualité de vie)

Ou même quand les deux composants nominaux sont concrets :

σοκολάτα αμυγδάλου
(chocolat aux amandes)

Ou encore quand le nom-tête est abstrait et le deuxième composant nominal est concret :

παρεκτροπή πυξίδας
(déviation du compas)

Nous pourrions donc conclure que le déterminant zéro ne dépend exclusivement ni des variantes savantes ni de la nature sémantique des composants nominaux.

Notons enfin que, quand le déterminant du deuxième complément nominal est l'article défini générique, il peut souvent être remplacé par le déterminant zéro, puisque, en grec moderne, la notion de genericité est aussi exprimée par le déterminant zéro. En guise d'illustration, citons les exemples suivants :

γλυκό (Ε + του) ταψιού
(gâteau du plateau)

δαχτυλίδι (Ε + των) αρραβώνων
(bague de fiançailles)

λάδι (Ε + της) ελιάς
(huile d'olive)

QUATRIEME PARTIE : DESCRIPTION LINGUISTIQUE

Chapitre 8. Etude syntaxico-sémantique des $N(E+DET:G)N:G$

8.1 Introduction

Dans cette partie, nous nous intéressons à étudier certaines propriétés syntaxiques et sémantiques des $N(E+DET:G)N:G$ recensés dans le cadre de cette étude. Plus particulièrement, nous nous focalisons sur l'étude des fonctions syntaxiques et sémantiques du génitif, la distinction entre $N(E+DET:G)N:G$ prédictifs et non-prédictifs et sur la classification sémantique des $N(E+DET:G)N:G$ et des $N1(E+DET:G)A:G N2:G$ à l'aide de traits sémantiques.

Dans un premier temps, nous analysons le complément de nom au génitif : « le génitif est en effet un cas problématique dans la mesure où plusieurs fonctions syntaxiques peuvent lui être attachées » (A. Fotopoulou 1993 : 260). Ainsi, après une brève présentation de la classification des génitifs proposée par les grammaires traditionnelles (A. Tzartanos 1946 : 107-116), nous étudions la formation du complément de nom au génitif sur la base de dérivations syntaxiques régulières et reproductibles et nous constatons leur blocage dans les noms composés $N(E+DET:G)N:G$.

Dans un deuxième temps, nous faisons la distinction entre noms prédictifs et non-prédictifs de structure $N(E+DET:G)N:G$. Cette tâche est d'une importance primordiale pour notre travail qui est tourné vers les applications du TAL, dans la mesure où notre principe d'avoir comme point de départ la phrase et pas le mot détermine la forme que prendront les dictionnaires électroniques : il faut donc séparer les opérateurs des arguments. Les informations linguistiques qu'il convient de donner ne sont pas de même nature selon qu'il s'agisse de prédicats ou d'arguments. Les prédicats peuvent avoir aussi une forme morphologique nominale. Il faut donc préciser la nature de leurs arguments (cf. G. Gross 1995a : 161-162). En d'autres termes, les noms prédictifs (simples ou composés) doivent figurer dans le dictionnaire électronique morphologique, mais ils doivent également être décrits en fonction des verbes supports avec qui ils se combinent et être représentés dans les tables du lexique-grammaire correspondantes. En revanche, les noms non-prédictifs (simples ou composés) doivent figurer uniquement dans les dictionnaires électroniques morphologiques.

Le dernier chapitre de cette partie porte sur la classification sémantique des $N(E+DET:G)N:G$. Nous décrivons certaines propriétés sémantiques qui ont été codées dans les tables du lexique-grammaire au moyen des traits sémantiques élémentaires comme « humain », « animé » ou « concret ». Cette description est nécessaire car, dans le cadre du lexique-grammaire, le lexique doit être structuré de façon cohérente avec la description des arguments dans le lexique-grammaire.

8.2 Le génitif dans les $N(E+DET:G)N:G$

En ce qui concerne le grec moderne, plusieurs études ont été consacrées au génitif. A titre indicatif, citons ici les travaux de A. Tzartanos 1946, I. Tsamadou 1984 et de A. Fotopoulou 1993c.

8.2.1 Le système casuel du grec moderne

L'étude des cas a intéressé presque tous les grammairiens et plusieurs travaux ont été consacrés aux systèmes casuels (L. Hjelmslev 1935, R. Jakobson 1936, N. Chomsky 1981).

Rappelons que le grec moderne est une langue à quatre cas : nominatif, génitif, accusatif, vocatif. Le cas morphologique du datif du grec ancien a depuis longtemps disparu du système casuel du grec moderne : on ne le retrouve actuellement que dans des structures figées (e.g. *λόγω*/en raison de, *ιδίοις όμμασι*/de ses propres yeux) (cf. aussi A. Fotopoulou 1993a, S. Voyatzi 2006). En grec moderne, le cas morphologique du datif du grec ancien a été remplacé par le cas morphologique du génitif (appelé depuis génitif-datif), par des groupes prépositionnels et par le cas morphologique de l'accusatif. Le vocatif est le cas de l'interpellation directe d'une personne (ou d'une chose) et, en tant que tel, il ne remplit pas de fonction « grammaticale ». Comme le note A. Fotopoulou (1993c : 260), « les trois autres cas sont essentiellement grammaticaux, d'autant plus qu'ils n'ont pas toujours, selon les flexions, des distinctions morphologiques ; c'est souvent l'article et la position dans la phrase qui indiquent le cas ».

I. Tsamadou (1984 : 3) signale que « la variation morphologique du génitif ne constitue qu'un résidu historique de l'ancienne diglossie grecque fondée sur la distinction entre la *katharevousa* (langue savante) et la *dimotiki* (langue démotique, courante) ». La même auteur remarque que « la variation morphosyntaxique constitue un phénomène extrêmement

complexe : le génitif s'affaiblit, et, à l'exception de quelques emplois productifs, est généralement remplacé par l'accusatif, par des groupes prépositionnels ou par des différentes structures phrastiques. La difficulté que certains noms présentent dans la formation du génitif singulier et pluriel ainsi que la difficulté de l'interprétation sémantique d'une structure, où plusieurs génitifs se succèdent, se répercutent sur la fréquence des occurrences du génitif » (I. Tsamadou 1984 : 3).

8.2.2 La classification du génitif adnominal en grec moderne

Dans le cadre de la présente étude, nous nous intéressons au génitif adnominal. Dans ce qui suit, nous présentons brièvement la classification des emplois adnominaux du génitif établie par A. Tzartanos (1946 : 107-116).

Le point de départ de A. Tzartanos (1946 : 107-116) est le cas morphologique du génitif. La classification qu'il propose est essentiellement fondée sur des critères sémantiques. Il distingue les six types de génitif adnominal suivants :

- i) le génitif possessif, qui reçoit différentes interprétations sémantiques, par exemple :

- le génitif possessif pur, qui désigne la possession

το βιβλίο του Νίκου/le livre le-Gms Nikos-Gms

(le livre de Nicolas)

- le génitif possessif qui désigne l'origine ou la parenté

ο παππούς του Νίκου/le grand-père le-Gms Nikos-Gms

(le grand-père de Nicolas)

- le génitif possessif qui désigne la dépendance

οι σύμμαχοι των Γάλλων/les alliées les-Gmp Français-Gmp

(les alliés des Français)

- le génitif possessif qui désigne le créateur d'une œuvre

η Οδύσσεια του Ομήρου/l'Odyssée le-Gms Homère-Gms

(l'Odyssée d'Homère)

- le génitif possessif qui désigne la provenance

η μυρωδιά του πεύκου/l'odeur le-Gns pin-Gns

(l'odeur du pin)

- le génitif possessif désignant le lieu

η μάχη του Μαραθώνα/la bataille le-Gms Marathon-Gms

(la bataille de Marathon)

- le génitif possessif désignant le temps

το όνειρο της αυγής/le rêve la-Gfs aube-Gfs

(le rêve de l'aube)

- le génitif possessif désignant l'usage ou le but

φόρεμα χορού/robe bal-Gms

(robe de bals)

- le génitif possessif qui désigne une comparaison

καρδιά παιδιού/cœur enfant-Gns

(cœur d'enfant)

- ii) le génitif qualificatif, qui reçoit également différentes interprétations sémantiques, par exemple

- le génitif qualificatif pur, qui désigne une qualité

κραυγή αγωνίας/cris angoisse-Gfs

(cri d'angoisse)

- le génitif qualificatif désignant l'extension spatiale ou temporelle

τακούνι δέκα πόντων/talon dix-Gmp centimètres-Gmp

(talon de 10 cm)

- le génitif qualitatif qui désigne l'âge

άνθρωπος εξήντα χρονών/personne soixante-Gmp ans-Gmp
(personne de soixante ans)

- le génitif qualitatif désignant le poids

γουρούνι εκατό οκάδων/porc cent-Gfp oques-Gfp
(un porc de cent oques)

- le génitif qualitatif désignant la valeur

κάλτσες εκατό δραχμών/chaussettes cent-Gfp drachmes-Gfp
(des chaussettes de cent drachmes)

iii) le génitif du contenu, qui reçoit différentes interprétations sémantiques, par exemple

- le génitif du contenu pur

αποθήκη σανού/entrepôt foin-Gms
(entrepôt de foin)

- le génitif explicatif ou descriptif

έγκλημα εσχάτης προδοσίας/crime ultime-Gfs trahison-Gfs
(crime de lèse majesté)

iv) le génitif partitif

τρεις στρατιώτες του πυροβολικού/trois soldats la-Gns artillerie-Gns
(trois soldats de l'artillerie)

v) le génitif subjectif et le génitif objectif

η διαδοχή των γεγονότων/la succession les-Gnp événements-Gnp
(la succession des événements)

η αγάπη του έθνους/l'amour la-Gns nation-Gns
(l'amour de la nation)

vi) le génitif désignant la cause

η χαρά της επιστροφής/la joie le-Gfs retour-Gfs
(la joie du retour)

Comme nous pouvons le constater, A. Tzartanos (1946 : 107-116) n'étudie pas les propriétés syntaxiques des différents types du génitif. De plus, il n'effectue aucune distinction entre les génitifs qui font partie de groupes nominaux libres et ceux qui font partie de groupes nominaux figés (noms composés). Rappelons que, dans le cadre de cette étude, nous entendons par groupe nominal libre de structure $N(E+DET:G)N:G$ les séquences dont les deux constituants, à savoir le nom-tête et le deuxième complément nominal au génitif, renvoient à deux référents différents, par exemple :

το σπίτι του Νίκου
(la maison de Nicolas)

En revanche, les deux constituants d'un nom composé de même structure forment une seule unité lexicale et sont donc indissociables (cf. I, 2.3), par exemple

ζώνη ασφαλείας
(ceinture de sécurité)

Nous nous proposons d'étudier la formation et le comportement du génitif adnominal sur la base de critères non seulement sémantiques, mais aussi sur la base de critères syntaxiques. Notre hypothèse est que la classification sémantique doit être principalement régie par des opérations syntaxiques régulières et reproductibles.

8.2.3 La formation du complément de nom au génitif en grec moderne

Comme nous l'avons signalé (cf. IV, 8.2.1), le grec moderne est une langue à cas. De ce fait, le complément de nom peut être soit prépositionnel (noté *Prép GN*) soit casuel (noté *GN:G*). Autrement dit, les compléments de nom peuvent être introduits par une préposition et, dans ce cas, le cas morphologique du groupe nominal en jeu dépend de cette préposition, comme par exemple :

N Prép GN =: άτομο με ειδικές ανάγκες/personne avec spéciaux besoins-Afp
(personne à mobilité réduite)

N Prép GN=: εισιτήριο μετ' επιστροφής/billet avec retour._{Gfs}
(billet aller-retour)

Dans le premier exemple, le groupe nominal introduit par la *Prép=: με/avec* est à l'accusatif alors que, dans le deuxième exemple, il est au génitif. Ce dernier est introduit par la *Prép=: μετά/avec*.

Mais, les compléments de nom peuvent également être introduits par la préposition zéro (*Prép=: E*) et, alors, le cas morphologique du groupe nominal non-prépositionnel est le plus souvent le génitif⁹⁷ :

N GN:G=: το βιβλίο του Νίκου/le livre le._{Gms} *Nikos*._{Gms}
(le livre de Nicolas)

N GN:G=: η δύση του ήλιου/le coucher le._{Gms} *soleil*._{Gms}
(le coucher du soleil)

Comme le note S. Voyatzi (2006 : 292), dans certains cas, on observe la possibilité de substitution du nom prépositionnel par le complément de nom casuel. Les deux compléments sont alors syntaxico-sémantiquement équivalents :

N Prép GN=: τα πόδια από την καρέκλα/les pieds._{Nnp} *de la chaise*._{Afs}
(les pieds de la chaise)

= *N GN:G=: τα πόδια της καρέκλας/les pieds*._{Nnp} *la chaise*._{Gfs}
(les pieds de la chaise)

En grec moderne, tout comme en français, la formation du complément de nom (qu'il soit prépositionnel ou casuel) est définie comme une opération en deux étapes : relativation d'une phrase à *Vsup*, puis réduction conjointe de la relative et du *Vsup* (cf. M. Gross 1981). A titre d'illustration, nous entreprenons cette analyse pour l'exemple suivant : *το βιβλίο του Νίκου/le livre de Nicolas*.

Ainsi, notre point de départ est la phrase élémentaire à *Vsup=: έχω/avoir*, qui est l'opérateur à lien du grec moderne (M. Gross 1981) :

⁹⁷ Certains linguistes distinguent, du point de vue terminologique, le « génitif propre » du « génitif datif », qui peut être remplacé par un complément prépositionnel en *σε/à GN* ou *από/de GN*. Sur ce point, cf. I. Tsamadou (1984), A. Fotopoulou (1990 ; 1993a, 1993b).

$[P] \# N_0 V_{sup} N_1 =:$

$[H \text{ Μαρία διαβάζει}] \# O \text{ Νίκος έχει ένα βιβλίο}$

$[La \text{ Maria lit}] \# Le \text{ Nikos a un-}_{Ans} \text{ livre-}_{Ans}$

Par relativation de la phrase à V_{sup} [*que* V_{sup}], nous obtenons :

$[P] \# N_1 \text{ που } V_{sup} N_0 =:$

$[H \text{ Μαρία διαβάζει}] \# Το \text{ βιβλίο που έχει ο Νίκος}$

$[La \text{ Maria lit}] \# Le-_{Nns} \text{ livre-}_{Nns} \text{ qu'a } le-_{Nms} \text{ Nikos-}_{Nms}$

La réduction à zéro de la relative à V_{sup} [*que* V_{sup} z.], fournit le complément de nom :

$[P] \# N_1 (GN_0:G) =:$

$[H \text{ Μαρία διαβάζει}] \# Το \text{ βιβλίο του Νίκου}$

$[La \text{ Maria lit}] \# Le-_{Nns} \text{ livre-}_{Nns} le-_{Gms} \text{ Nikos-}_{Gms}$

En appliquant le second membre nominal au premier membre phrastique (phrase principale), nous obtenons la phrase élémentaire ci-dessous :

$N_0 V N_1 (= N (GN:G)) =:$

$H \text{ Μαρία διαβάζει το βιβλίο του Νίκου}$

$La-_{Nfs} \text{ Maria-}_{Nfs} \text{ lit } le-_{Ans} \text{ livre-}_{Ans} le-_{Gms} \text{ Nikos-}_{Gms}$

Notons, encore, qu'une propriété caractéristique des compléments de nom au génitif est la pronominalisation sous la forme d'un possessif (simple ou composé) :

$N_0 V N_1 (= GN:G) =:$

$H \text{ Μαρία διαβάζει το βιβλίο του Νίκου}$

$La-_{Nfs} \text{ Maria-}_{Nfs} \text{ lit } le-_{Ans} \text{ livre-}_{Ans} le-_{Gms} \text{ Nikos-}_{Gms}$

(Maria lit **le livre de Nicolas**)

$N_0 V N_1 (= Dd\acute{e}f (N Poss_s + Poss_c N)) =:$

$H \text{ Μαρία διαβάζει το (βιβλίο του+δικό του βιβλίο)}$

La-Nfs Maria-Nfs lit le-Ans (livre-Ans à lui-Gms+sien-Ans à lui-Gms livre-Ans)
 (Maria lit son livre)

Toutefois, la possessivation du complément de nom présente généralement des contraintes lexicales qui dépendent de la nature syntaxico-sémantique du *N* (*N-hum*, dans l'exemple précité) (cf. I. Tsamadou 1984).

8.2.4 Blocage des dérivations syntaxiques régulières et reproductibles dans les *N (E+DET:G) N:G*

Comme ceci a été explicité dans le paragraphe précédent, la formation du complément de nom au génitif d'une structure *N (E+DET:G) N:G* résulte des dérivations syntaxiques régulières, alors celle-ci devrait être qualifiée d'unité lexicale libre (ou compositionnelle) ou, au moins, de nom composé dont le degré de figement n'est pas total (cf. I, 2.2.9). Cette analyse, fondée sur des critères purement syntaxiques, permet souvent d'assigner au complément de nom au génitif une étiquette sémantique comparable à celles proposées par A. Tzartanos (1946).

Dans le cas contraire où la formation du complément de nom au génitif d'une séquence nominale de structure *N (E+DET:G) N:G* n'obéit pas à ces dérivations syntaxiques régulières, alors celle-ci devrait être qualifiée de nom composé entièrement figé. Nous avons abordé les contraintes syntaxiques dans I, 2.2.3-2.2.5, il n'y pas donc besoin d'y revenir plus explicitement ici. Nous nous contentons de donner un exemple :

H María στραγγίζει τα μαλλιά αγγέλου
La-Nfs Maria-Nfs égoutte les-Anp cheveux-Anp ange-Gms
 (Marie égoutte les cheveux d'ange)

Aucun verbe support ne semble pouvoir être appliqué au mécanisme de la dérivation syntaxique du complément de nom au génitif *αγγέλου*/d'ange.

Chapitre 9. Distinction entre $N (E+DET:G) N:G$ prédicatifs et $N (E+DET:G) N:G$ non-prédicatifs

9.1 Délimitation de l'étude des noms prédicatifs

Dans le cadre de notre étude, nous nous sommes fixée l'objectif d'effectuer la distinction entre noms prédicatifs et noms non-prédicatifs au sein des noms composés de structure $N (E+DET:G) N:G$. Cette tâche est d'une importance primordiale pour notre travail dans la mesure où les informations linguistiques qu'il convient de donner ne sont pas de même nature selon qu'il s'agit de prédicats ou d'arguments. En d'autres termes, les noms prédicatifs (simples ou composés) ne doivent pas figurer uniquement dans le dictionnaire électronique morphologique ; ils doivent également être décrits en fonction des verbes supports avec qui ils se combinent et être représentés dans les tables du lexique-grammaire correspondantes. En revanche, en ce qui concerne les noms non-prédicatifs (simples ou composés), leur présence dans les dictionnaires électroniques morphologiques est suffisante.

Les propriétés des noms prédicatifs de structure $N (E+DET:G) N:G$ sont identiques à celles des noms prédicatifs simples. Nous considérons donc que les $N (E+DET:G) N:G$ prédicatifs du grec moderne doivent être inclus dans les tables du lexique-grammaire au même titre que les noms prédicatifs simples, comme ceci a été fait pour le français quoique dans de faibles proportions (cf. L. Danlos 1980, J. Giry-Schneider 1987, G. Gross 1990).

La question de distinguer les noms prédicatifs des noms non-prédicatifs est cruciale. Comme les noms prédicatifs et les noms non-prédicatifs n'ont pas le même comportement syntaxique, un certain nombre de propriétés syntaxiques propres aux noms prédicatifs permettent d'effectuer une distinction de manière formelle et de délimiter les noms prédicatifs. Nous rappelons les critères syntaxiques de distinction entre noms prédicatifs et noms non-prédicatifs dans la section IV, 9.3.1-9.3.4.

9.2 La notion de nom prédicatif et de verbe support

Toute phrase simple est articulée autour d'un noyau prédicatif qui sélectionne un ou plusieurs arguments : le nombre et le type des arguments sont déterminés par le prédicat. On définit donc les arguments comme les variables du prédicat sémantique. Ainsi, la phrase construite autour du prédicat $V = : \beta\lambda\acute{\epsilon}\pi\omega$ /regarder :

Η Μαρία βλέπει μία ταινία

Maria regarde un film

peut être représentée de la manière suivante :

βλέπω (Μαρία, ταινία)

(regarder (Maria, film))

où le $V =$: *βλέπω*/regarder est le prédicat, *Μαρία*/Maria et *ταινία*/film sont les arguments de ce prédicat. Les arguments d'un prédicat sémantique, comme par exemple un verbe, occupent des positions syntaxiques (sujet, premier complément, deuxième complément etc.) en tant qu'actants de ce verbe. Par exemple, pour le prédicat précité : *βλέπω*/regarder (x , y), l'argument x correspond au sujet du verbe (*Μαρία*/Maria) et l'argument y à son complément (*ταινία*/film) (cf. A.-V. Pantazara 2003 :11).

Dans l'optique transformationnelle harrissienne, le prédicat ne s'identifie pas à une seule catégorie morphologique. Il est susceptible de réalisations multiples. Les mots prédictifs peuvent appartenir à différentes classes, comme par exemple des verbes, des noms ou des adjectifs.

L'étude des noms prédictifs est associée à celles des « verbes supports ». La notion du « verbe support » a été introduite en 1964 par Z. S. Harris pour désigner des verbes qui ne sont pas porteurs de sens et qui ne sélectionnent pas les actants nominaux de la phrase. Ils peuvent être supports de nominalisations. Par exemple :

Η Μαρία μελέτησε τα ρήματα

Maria a étudié les verbes

Η Μαρία έκανε τη μελέτη των ρημάτων

Maria a fait l'étude des verbes

Cependant, il existe des cas où le nom prédictif n'est associé à aucune forme verbale. Dans ce cas, il s'agit d'un prédicat nominal dit autonome, isolé et non dérivé. Voici un exemple :

Η Μαρία έκανε μία γκάφα

Maria a fait une gaffe

Dans les exemples précités, la forme verbale *κάνω*/faire n'a aucun rôle prédicatif. Ce rôle est tenu par les noms prédicatifs *μελέτη*/étude et *γκάφα*/gaffe auxquels ces verbes sont associés. Les verbes supports sont des auxiliaires qui portent les marques de temps-aspect et de personne-nombre (cf. A.-V. Pantazara 2003:12).

Nous rappelons ici les propriétés principales des verbes supports telles qu'elles sont exposées dans G. Gross 1996b et telles qu'elles sont reprises dans A.-V. Pantazara 2003 : 12 :

« i) Le verbe support est toujours sélectionné par le nom prédicatif ou l'adjectif avec lequel il apparaît. La sélection du support constitue une des propriétés distributionnelles du nom prédicatif ou de l'adjectif prédicatif. Le support des adjectifs prédicatifs est par excellence le verbe support *είμαι*/être, alors que les noms prédicatifs sélectionnent divers verbes supports selon leur nature sémantique.

ii) Le verbe support peut être effacé dans une phrase, sans que le nom prédicatif ou l'adjectif, auquel le verbe support est associé, perde son statut prédicatif. En revanche, l'effacement d'un verbe prédicatif supprime *de facto* la phrase.

iii) Les verbes supports ne peuvent pas faire l'objet d'un changement de catégorie, ils n'ont pas de forme nominale ou adjectivale ; les transformations morphologiques (nominalisation, adjectivation) sont le fait des prédicats.

iv) La fonction de nominalisation n'est pas pour les verbes supports une propriété définitionnelle car une telle situation n'est pas, et de loin, la plus courante : le nombre des prédicats autonomes est supérieur à celui des déverbaux. L'existence d'un verbe associé est un accident morphologique de ce point de vue. »

9.3 Les critères syntaxiques de distinction entre noms prédicatifs et noms non-prédicatifs

Dans ce sous-chapitre, nous reprenons les critères de distinction déjà présentés par plusieurs auteurs (cf. notamment J. Giry-Schneider 1987 et A. Monceaux 1993 pour le français, A. Fotopoulou 1993a, A.-V. Pantazara 2003 et V. Sfetsiou 2002 ; 2007 pour le grec moderne). C'est par ces critères qu'est explicitée l'intuition que des substantifs comme *ανταλλαγή πληθυσμών*/échange de populations ou *πόλεμος νεύρων*/guerre des nerfs, (associables au verbe support *κάνω*/faire), et *γκολ της νίκης*/but de la victoire, (associable au verbe support *βάζω*/marquer), désignent des procès, alors que *καφές φίλτρου*/café filtre et *γυαλιά*

ηλίου/lunettes de soleil, (également associables à *κάνω*/faire et *βάζω*/mettre (*Η Μαρία κάνει έναν καφέ φίλτρου*/Maria fait un café filtre, *Η Μαρία έβαλε τα γυαλιά ηλίου*/Maria a mis les lunettes de soleil)) désignent des objets concrets (cf. J. Giry-Schneider 1987 : 27).

Les propriétés syntaxiques qui permettent d'effectuer la distinction entre noms prédicatifs et noms non-prédicatifs sont les suivantes :

- La relation entre le sujet sémantique du nom prédicatif et le sujet grammatical du verbe support
- L'existence d'un groupe nominal indépendant qui soit synonyme de la phrase à verbe support et qui conserve le complément *Prép NI*
- La question par *τι;/que ?*, souvent douteuse
- La « descente » ou « montée » de l'adverbe sous forme de modifieur adjectival du nom prédicatif.

Nous allons par la suite tester ces propriétés sur deux $N (E+DET:G) N:G$: d'une part le nom prédicatif *άλμα θανάτου*/saut de mort, et d'autre part le nom non-prédicatif *καφές φίλτρου*/café filtre.

9.3.1 Le sujet sémantique du nom prédicatif est le sujet grammatical du verbe support

Considérons les phrases suivantes :

Ο Νίκος έκανε ένα άλμα θανάτου
(Nicolas a fait un saut de la mort)

Ο Νίκος έκανε έναν καφέ φίλτρου
(Nicolas a fait un café filtre)

Dans le premier cas, le sujet grammatical ($N0$) du verbe support *έκανε*/a fait est en même temps le sujet sémantique du nom prédicatif *άλμα θανάτου*/saut de mort. Cette relation de coréférence rend compte de l'interdiction d'un complément de nom au génitif, source d'un pronom possessif à droite du nom prédicatif. Ainsi, dans le premier cas, le nom prédicatif *άλμα θανάτου*/saut de mort n'admet pas de complément de nom au génitif :

**Ο Νίκος έκανε το άλμα θανάτου (της Μαρίας + της)*

(*Nicolas a fait le saut de mort de Maria)

En revanche, ceci est tout à fait possible dans le deuxième cas. Il s'agit ici d'une construction avec *κάνω*/faire, qui peut accepter un complément de nom au génitif ou un pronom possessif non-coréférent au sujet de la phrase :

Ο Νίκος έκανε τον καφέ φίλτρου της Μαρίας

(Nicolas a fait le café filtre de Maria)

?Ο Νίκος έκανε τον καφέ φίλτρου της

?Nicolas a fait le café filtre-*Gns* elle-*Gfs*

(Nicolas a fait son café filtre)

La relation du *N₀* par rapport au nom prédicatif distingue les phrases à verbe support des phrases qu'un nom prédicatif peut former avec d'autres verbes prédicatifs. Ainsi, le nom prédicatif *άλμα θανάτου*/saut de mort peut apparaître dans une phrase avec le verbe *περιγράφω*/décrire :

Ο Νίκος περιέγραψε το άλμα θανάτου

(Nicolas a décrit le saut de mort)

Pour cette phrase on n'observe pas les mêmes contraintes concernant l'interdiction d'un complément de nom au génitif, source d'un pronom possessif à droite du nom prédicatif. On peut donc avoir :

Ο Νίκος περιέγραψε το άλμα θανάτου της Μαρίας

(Nicolas a décrit le saut de mort de Maria)

Ce contraste met en évidence le fait qu'un nom prédicatif a un sujet. Dans une construction à verbe support, ce sujet a la fonction grammaticale de *N₀*, il constitue donc en même temps le sujet de la phrase. Par contre, dans le cas d'une phrase à verbe prédicatif, comme *περιγράφω*/décrire, la coréférence entre le sujet de la phrase et le sujet du nom prédicatif n'est pas obligatoire.

Enfin, la phrase :

Ο Νίκος περιέγραψε το άλμα θανάτου

(Nicolas a décrit le saut de mort)

sera analysée comme phrase complexe qui met en jeu deux phrases simples, dont l'une est une phrase à verbe support *κάνω*/faire. La phrase complexe est obtenue par réduction du verbe support de la phrase simple à nom prédicatif :

Ο Νίκος περιέγραψε τούτο # (Ο Νίκος + Η Μαρία) κάνει άλμα θανάτου

(Nicolas a décrit ceci # (Nicolas + Maria) fait un saut de la mort)

Comme le note A.-V. Pantazara (2003 : 29), « la possibilité ou non d'insérer un complément de nom au génitif (source d'un pronom possessif) à la suite d'un nom est un excellent critère pour détecter si ce nom est un nom prédicatif ou non ».

9.3.2 Existence d'un groupe nominal indépendant qui soit synonyme de la phrase à verbe support et qui conserve le complément *Prép N_I*

Selon J. Giry-Schneider (1987: 28), « on peut former un groupe nominal de la forme *Le N de N₀ Prép N_I* à partir d'une forme à verbe support par réduction de la relative, selon la règle [Réd *Vsup*] (M. Gross 1981) ». A titre indicatif, citons les exemples suivants :

Η Μαρία έκανε ένα άλμα θανάτου στο στάδιο

Maria a fait un saut de la mort au stade

Το άλμα θανάτου της Μαρίας στο στάδιο [ήταν εντυπωσιακό]

Le saut de mort de Maria au stade [était impressionnant]

= *Το άλμα θανάτου που η Μαρία έκανε στο στάδιο [ήταν εντυπωσιακό]*

= Le saut de mort que Maria a fait au stade [était impressionnant]

Ce qui caractérise les groupes nominaux ainsi obtenus, c'est qu'ils sont synonymes sans ambiguïté de la forme de base à verbe support et que le complément *Prép N_I* s'y trouve conservé, ce qui n'est pas le cas avec les noms concrets comme *καφές φίλτρου*/café filtre :

Ο Νίκος έκανε έναν καφέ φίλτρου στη Μαρία

Nicolas a fait un café filtre à Maria

Ο καφές φίλτρου που ο Νίκος έκανε στη Μαρία [ήταν δυνατός]

Le café filtre que Nicolas a fait à Maria [était fort]

**Ο καφές φίλτρου του Νίκου στη Μαρία [ήταν δυνατός]*

**Le café filtre de Nicolas à Maria [était fort]*

Le seul groupe nominal possible est le groupes nominal suivant :

GN = : Le N de N0 = ο καφές φίλτρου του Νίκου/le café filtre de Nicolas

Mais ce groupe nominal sans complément prépositionnel est très ambigu, car il peut se rattacher à toutes sortes de constructions verbales : *ο καφές φίλτρου που (έκανε + ήπιε) ο Νίκος/le café filtre que Nicolas (a fait + a bu) etc.*

En revanche, il n'y a aucune ambiguïté dans le groupe nominal *το άλμα θανάτου της Μαρίας στο στάδιο/le saut de mort de Maria au stade*, qui peut s'interpréter uniquement comme *το άλμα θανάτου που η Μαρία έκανε στο στάδιο/le saut de mort que Maria a fait au stade*.

9.3.3 La question par *τι;/que* ?, souvent douteuse

On observe que la question par *τι;/que* ? donne des paires question-réponse acceptables pour les noms concrets, comme par exemple :

Ο Νίκος κάνει έναν καφέ φίλτρου (E + στην κουζίνα)

- *Τι κάνει ο Νίκος (E + στην κουζίνα);*
- *Έναν καφέ φίλτρου*

(Nicolas fait un café filtre (E + dans la cuisine)

- *Que fait Nicolas (E + dans la cuisine) ?*
- *Un café filtre)*

En revanche, la question par *τι;/que* ? donne souvent des paires question-réponse inacceptables avec les verbes supports. Par exemple :

Η Μαρία έκανε ένα λάθος απροσεξίας στους υπολογισμούς της

- * - Τι έκανε η Μαρία στους υπολογισμούς της;*
- *Ένα λάθος απροσεξίας*

(Maria a fait une erreur d'inattention dans ses calculs

- * - Qu'est-ce que Maria a fait dans ses calculs ?*
- *Une erreur d'inattention)*

Cependant, à côté des énoncés inacceptables précités, on peut trouver des exemples qui se prêtent à la question par *τι;/que* ?. Par exemple :

Ο Νίκος κάνει μία μελέτη σκοπιμότητας

- *Τι κάνει ο Νίκος;*
- *Μία μελέτη σκοπιμότητας*

(Nicolas fait une étude de faisabilité)

- *Que fait Nicolas ?*
- *Une étude de faisabilité)*

Ces différences d'acceptabilité s'expliquent par l'extrême ambiguïté de la question par *τι;/que* ?. En l'absence de tout complément prépositionnel, on peut répondre à la question par *τι;/que* ? soit par un verbe quelconque soit par le verbe support lui-même. Par exemple :

- *Τι κάνει ο Νίκος;*
- *(Σκέφτεται + κάνει μία μελέτη σκοπιμότητας)*
- *Que fait Nicolas ?*
- *(Il réfléchit + il fait une étude de faisabilité)*

Dans ce cas donc, le verbe *κάνω*/faire de la question n'est pas le *Vsup* = : *κάνω*/faire.

On observe que dans le cadre de la question par *τι;/que* ?, on ne peut pas dissocier certains compléments *Prép NI* du nom dont ils sont arguments (*λάθος απροσεξίας στους υπολογισμούς*/erreur d'inattention dans les calculs). Cependant, cette contrainte ne s'observe pas quand *Prép NI* est un locatif (J. Giry-Schneider 1987: 30). Par exemple :

Η Μαρία έκανε ένα άλμα θανάτου στο στάδιο

- *Τι έκανε η Μαρία στο στάδιο;*
- *Ένα άλμα θανάτου*

(Maria a fait un saut de la mort au stade)

- *Qu'est-ce que Maria a fait au stade ?*
- *Un saut de la mort)*

Η Μαρία έκανε ένα ταξίδι αναψυχής στην Ιταλία

- *Τι έκανε η Μαρία στην Ιταλία;*

- *Ένα ταξίδι αναψυχής*

(Maria a fait un voyage d'agrément en Italie)

- Qu'est-ce que Maria a fait en Italie ?
- Un voyage d'agrément)

Comme en conclut J. Giry-Schneider (1987: 30), « ces faits ne sont pas aisés à interpréter ; si l'on parle en termes de groupe nominal, on dira que la tête de ces *GN* tantôt s'extrait, tantôt ne s'extrait pas, ce qui remet en question l'idée souvent émise que la tête d'un *GN* ne s'extrait pas ; on constate en fait que l'extraction de cette tête dépend du complément prépositionnel qui suit et du verbe principal, et qu'il serait vain de chercher des généralisations en l'absence d'autres données sur ce sujet ».

9.3.4 Descente ou montée de l'adverbe

Il s'agit de l'emploi équivalent que peuvent avoir un adverbe et un adjectif dans une phrase à verbe support. Soit les paires *ανόητος/stupide-ανόητα/stupidement* et *συχνός/fréquent-συχνά/fréquemment*, on aura avec *λάθος απροσεξίας/erreur d'inattention* les phrases synonymes suivantes :

Η Μαρία έκανε ανόητα ένα λάθος απροσεξίας

(Maria a fait stupidement une erreur d'inattention)

= *Η Μαρία έκανε ένα ανόητο λάθος απροσεξίας*

(= Maria a fait une erreur d'inattention stupide)

Η Μαρία κάνει συχνά λάθη απροσεξίας

(Maria fait fréquemment des erreurs d'inattention)

= *Η Μαρία κάνει συχνά λάθη απροσεξίας*

(= Maria fait des erreurs d'inattention fréquentes)

Avec les noms non-prédicatifs *γλυκό του κουταλιού/gâteau de cuillère* et *καφές φίλτρου/café filtre*, seul l'adverbe est possible :

Η Μαρία κάνει συχνά καφέδες φίλτρου

Maria fait fréquemment des cafés filtre

**Η Μαρία κάνει συχνούς καφέδες φίλτρου*

**Maria fait des cafés filtre fréquents*

Selon J. Giry-Schneider (1987: 31), « cette équivalence apparaît souvent avec des adjectifs de sens psychologique comme stupide, mais aussi avec des adverbes à sens aspectuel comme continuels ou fréquents. De tels adjectifs aspectuels sont plus compatibles avec des noms supposés prédictifs qu'avec les noms concrets ».

L'ensemble de ces critères nous a permis d'effectuer la distinction entre les $N (E+DET:G) N:G$ prédictifs et les $N (E+DET:G) N:G$ non-prédictifs. Nous avons dressé une liste de 534 noms prédictifs de structure $N (E+DET:G) N:G$. A titre indicatif, citons les exemples suivants :

απεργία πείνας (κάνω)

grève de la faim (faire)

απόπειρα αυτοκτονίας (κάνω)

tentative de suicide (faire)

άφεση αμαρτιών (δίνω)

absolution (donner)

γκολ της νίκης (βάζω)

but de la victoire (marquer)

διακίνηση όπλων (κάνω)

trafic d'armes (faire)

θέση μάχης (παίρνω)

position de combat (prendre)

πλύση εγκεφάλου (κάνω)

lavage de cerveau (faire)

Nous proposons l'ajout du trait **Npréd** dans le dictionnaire électronique morphologique pour les entrées qui constituent des noms prédictifs (cf. V, 11.3.1.1). A l'heure actuelle, nous avons effectué ce codage uniquement dans les tables du lexique-grammaire (cf. V, 11.3.3). La

description syntaxique des $N(E+DET:G)N:G$ prédictifs à l'aide des tables du lexique-grammaire reste à faire.

9.4 Problèmes de distinction entre expression à éléments figés et expression à verbe support

Avant de conclure nos remarques sur les $N(E+DET:G)N:G$, nous voudrions aborder le sujet souvent délicat de distinction entre expression à éléments figés et expression à verbe support. Dans le cadre de cette étude, il nous a fallu effectuer cette distinction afin de délimiter les entrées qui devaient être codées comme *Npréd* dans nos ressources lexicales, mais nous n'avons pas effectué une analyse approfondie de ce phénomène.

Selon A. Fotopoulou (1993a :7), « les phrases figées se définissent par le fait qu'un ou plusieurs actants sont lexicalement invariables par rapport au verbe ». Par exemple, dans les phrases :

Η Μαρία πήρε το βάπτισμα του πυρός χθες

(Maria a reçu le baptême du feu hier)

[Η Μαρία έμεινε στήλη άλατος] όταν άκουσε τα νέα

[La Maria est restée stèle-*Afs* sel-*Gns*] quand elle a entendu les nouvelles

([Maria est restée bouche bée] quand elle a su les nouvelles)

c'est la combinaison invariable respectivement entre *πήρε/a* reçu et *βάπτισμα του πυρός/baptême du feu* et entre *έμεινε/est restée* et *στήλη άλατος/stèle sel-*Gns** qui les rend figées.

« Outre ce critère de forme, des intuitions de sens interviennent : le sens des mots ne permet pas d'interpréter leur combinaison » (A. Fotopoulou 1993 : 7). Le sens des deux phrases précitées n'est donc pas calculable à partir du sens de chacun des mots qui les constituent.

Mais la distinction entre expression à éléments figés et expression à verbe support (cf. IV, 9.2) est souvent délicate. Du point de vue lexical, les expressions *Vsup N* constituent des combinaisons spécifiques. Faut-il les intégrer à la liste des expressions à éléments figés alors qu'elles ont des sujets et des compléments lexicalement libres (comme *η Μαρία δεν έδωσε σημεία ζωής στο Νίκο/Maria n'a pas donné signe de vie à Nicolas*) par opposition aux formes telles que *Η Μαρία πήρε το βάπτισμα του πυρός/Maria a reçu le baptême du feu* ? « S'il suffit

qu'un seul terme soit figé pour que l'on parle d'expression figée alors *faire feu sur quelqu'un* fera partie de ces expressions ; par contre, dans une phrase comme *Paul fait un câlin à Marie*, aucun élément n'est figé ; ce qui demeure c'est la spécificité d'une telle phrase, spécificité des phrases à *Vsup* » (J. Giry-Scheider 1987: 93).

Comme le remarque J. Giry-Schneider (1987: 87), cette confusion entre expressions à éléments figés et expressions à verbes support est favorisée par les dictionnaires usuels qui appellent souvent « locutions » les expressions à verbe support. Elle est également observée chez certains linguistes qui ont choisi comme critère de figement l'absence de déterminant, critère traditionnel servant à définir les expressions à éléments figés (cf. J. Giry-Schneider 1987: 87).

Chapitre 10. Classification sémantique des $N(E + DET:G)N:G$ et des $N1(E+DET:G)A:GN2:G$

10.1 Introduction

Ce chapitre porte sur des questions sémantiques. Nous allons décrire quelques propriétés sémantiques que nous avons étudiées et codées dans les tables du lexique-grammaire⁹⁸ afin d'établir une classification sémantique des noms composés binaires de structure $N(E + DET:G)N:G$ et des noms composés complexes (ou surcomposés) de structure $N1(E+DET:G)A:GN2:G$. Nous avons utilisé pour cela des traits sémantiques élémentaires tels que « concret », « animé » ou « humain ». La présentation et la discussion de cette classification font l'objet de ce chapitre.

« Dans un dictionnaire électronique, l'information sémantique n'est pas utilisée directement par un humain, mais elle est destinée à des algorithmes. Par exemple, entre humains, il est inutile de spécifier que *automobile* est un **objet concret**. Par contre, cette indication est indispensable pour la cohérence du système des marques » (M. Gross 1995 : 10).

Comme le signale M. Gross (1995 : 10), « le codage d'un lexique électronique d'une grande taille pose des problèmes particuliers. Non seulement le système sémantique doit être cohérent, mais à un niveau pratique, la cohérence de chaque marque doit être assurée pour le dictionnaire complet ».

Selon A. Monceaux (1993: 184), « la description des propriétés sémantiques des noms du lexique est nécessaire : 1) parce qu'elle permet un traitement plus uniforme des homographes ou des variantes dans le dictionnaire et en facilite ainsi la consultation et la mise à jour par des lecteurs humains ; mais également : 2) parce qu'un programme d'analyse utilisant le lexique-grammaire doit avoir accès à un lexique structuré sémantiquement de façon cohérente avec la description des arguments dans le lexique-grammaire ».

⁹⁸ Précisons que les traits sémantiques doivent figurer également dans le dictionnaire électronique morphologique. L'élaboration du programme qui permet l'ajout automatique de ces traits dans le dictionnaire électronique morphologique est en cours.

10.2 Les traits sémantiques élémentaires

« On doit admettre la nécessité d'évoquer, même approximativement, la distribution des arguments des prédicats lorsque l'on veut élaborer un dictionnaire syntaxique : des traits syntaxico-sémantiques, même très généraux, sont utilisés dans le but de désambiguïser différents sens (ou emplois) d'une même forme, verbe, adjectif ou nom » A. Monceaux (1993: 185). Par exemple, le trait « humain » suffit pour différencier deux verbes *προσβάλλω*/insulter + atteindre sur le critère de leurs distributions :

	<i>Nhum προσβάλλω/insulter Nhum</i>
= :	<i>Ο Νίκος πρόσβαλε τη Μαρία</i> Nicolas a insulté Maria
	<i>N-hum προσβάλλω/atteindre Nhum</i>
= :	<i>Ο καρκίνος προσβάλλει ανθρώπους κάθε ηλικίας</i> Le cancer atteint des personnes de tout âge

Les traits sémantiques généraux suivants peuvent être utilisés pour la classification des arguments des prédicats :

$N_i =: Nhum$	pour les substantifs humains, e.g.	<i>παιδί</i> /enfant
$N_i =: Nanim$	pour les substantifs animés non-humains, e.g.	<i>σκύλος</i> /chien
$N_i =: Nconc$	pour les substantifs concrets, e.g.	<i>σπίτι</i> /maison
$N_i =: Nabs$	pour les substantifs abstraits, e.g.	<i>κατάχρηση</i> /abus
$N_i =: Ncoll$	pour les substantifs collectifs, e.g.	<i>σχολείο</i> /école

Notons que les notations $N_i =: Nhum$ et $N_i =: Nconc$ correspondent à des propriétés formelles des arguments du prédicat, alors que la notation $N_i =: Nabs$ ne correspond pas à des propriétés formelles ; elle répond à une intuition sémantique et regroupe les noms prédictifs dans une classe résiduelle (cf. IV, 10.7). Un critère formel qu'on pourrait admettre pour les $N_i =: Nabs$ est l'absence de notations $N_i =: Nhum$ et $N_i =: Nconc$.

Les traits sémantiques utilisés dans le cadre du lexique-grammaire marquent des propriétés du prédicat et permettent la désambiguïsation des entrées syntaxiques sur le critère de leurs propriétés distributionnelles.

Dans IV, 10.4-10.6, nous discutons de problèmes posés par une description du lexique au moyen des traits sémantiques humain, animé (non humain), concret et abstrait. Signalons qu'il existe un problème important, celui des emplois métonymiques ou métaphoriques des substantifs, dont certains sont lexicalisés et enregistrés dans les dictionnaires (cf. A. Monceaux 1993 : 187). De manière générale, les substantifs sont susceptibles d'interprétations contextuelles différentes.

10.3 Les conventions du codage

10.3.1 Traits binaires

Dans le cadre de cette étude, nous avons utilisé des traits sémantiques binaires, ce qui signifie qu'un nom qui n'a pas été codé « concret » doit nécessairement être considéré comme « non-concret » etc. Etant donné que plusieurs noms de cette étude sont ambigus, nous avons été contrainte à dédoubler les entrées. Précisons ici que nous avons utilisé les traits sémantiques binaires suivants : Nhum, Nhumcoll, Nanim, Nconc et Nabs.

10.3.2 Traits exclusifs

Les traits dont nous nous sommes servie sont des traits exclusifs, c'est-à-dire des traits qui déterminent une et une seule interprétation du nom. Précisons qu'un nom peut recevoir plusieurs traits exclusifs. Par exemple, les noms comme *άσυλο απόρων*/maison de repos, *άσυλο γερόντων*/maison de retraite, *άσυλο τυφλών*/institut d'aveugles peuvent recevoir deux traits sémantiques exclusifs en fonction de leur interprétation : celui de « concret » quand ils désignent les bâtiments respectifs, mais aussi celui de « humain collectif » lorsqu'ils désignent des « tenants humains », c'est-à-dire les gens qui y sont logés.

De la même façon, les noms comme *αρνάκι (Ε+του) γάλακτος*/agneau de lait, *γουρουνάκι (Ε+του) γάλακτος*/porc de lait etc. s'interprètent soit comme des « animés » lorsqu'ils désignent les petits animaux respectifs, soit comme « concrets » lorsqu'ils désignent le type de viande que l'on mange.

Comme le note A. Monceaux 1993 : 192, « le dédoublement des entrées pourrait être utilisé pour distinguer les cas de réelles homonymies pour lesquels l'interprétation contextuelle du nom n'était pas prévisible, des cas où les différentes interprétations contextuelles du nom sont en partie prévisibles, comme c'est le cas des noms désignant des lieux "contenants d'humains" ».

Pour les exemples ci-dessous, on pourrait donc avoir le codage suivant :

άσυλο απόρων/maison de repos (1) *Nconc*

άσυλο απόρων/maison de repos (2) *Nhumcoll*

άσυλο γερόντων/maison de retraite (1) *Nconc*

άσυλο γερόντων/maison de retraite (2) *Nhumcoll*

άσυλο τυφλών/institut d'aveugles (1) *Nconc*

άσυλο τυφλών/institut d'aveugles (2) *Nhumcoll*

αρνάκι (E+του) γάλακτος/agneau de lait (1) *Nconc*

αρνάκι (E+του) γάλακτος/agneau de lait (2) *Nanim*

γουρουνάκι (E+του) γάλακτος/porc de lait (1) *Nconc*

γουρουνάκι (E+του) γάλακτος/porc de lait (2) *Nanim*

Nous avons codé les propriétés sémantiques des *N (E+DET:G) N:G* et des *N1 (E+DET:G) A:G N2:G* dans les tables respectives du lexique-grammaire. A titre indicatif, citons un extrait issu de notre table du lexique-grammaire pour les *N (E+DET:G) N:G* :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
N	DET:G	N:G	N DET:G N:G	N:G	DET:G=E	DET:G N:G	N:G	N DET:G N:G=N:G-a N		N:G-a	N DET:G N:G=Ns	Ns	N DET:G N:G=Npréd	N DET:G N:G=Nhum	N DET:G N:G=Nhumcoll	N DET:G N:G=Nanim	N DET:G N:G=Nconc	N DET:G N:G=Nabs	N DET:G N:G=TS
άσυλο	-	ανιάτων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	+	-	-	-	-
άσυλο	-	ανιάτων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	-	-	+	-	-
άσυλο	-	απόρων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	+	-	-	-	-
άσυλο	-	απόρων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	-	-	+	-	-
άσυλο	-	γερόντων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	+	-	-	-	-
άσυλο	-	γερόντων	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	-	-	+	-	-
άσυλο	-	τυφλών	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	+	-	-	-	-
άσυλο	-	τυφλών	-	+	-	-	-	-	ΑΕ>	-	ΑΕ>	-	-	-	-	-	+	-	-

Figure 2 : Extrait de la table du lexique grammairal pour les $N (E+DET:G) N:G$

Nous voyons dans cet extrait que chacune des entrées *ἀσυλο ἀπόρων*/maison de repos etc. est codée deux fois, une fois comme Nhumcoll (nous avons marqué « + » dans la colonne O) et une fois comme Nconc (nous avons marqué « + » dans la colonne Q).

10.4 Les $N (E+DET:G) N:G$ et les $NI (E+DET:G) A:G N2:G$ humains

10.4.1 La notion sémantique « humain »

« La notion de nom « humain » correspond d'abord à une intuition sémantique : les noms référant à une personne ou un ensemble de personnes sont des noms « humains », qu'on dira respectivement « individuel » (notés Nhum) ou « collectifs » (notés Nhumcoll) » (A. Monceaux 1993 : 192-193). A titre indicatif, citons les exemples suivants de $N (E+DET:G) N:G$ et $NI (E+DET:G) A:G N2:G$ humains et humains collectifs :

Nhum

ανάπηρος πολέμου/mutilé de guerre

αντιρρησίας συνειδήσεως/objecteur de conscience

γυνή της απωλείας/femme corrompue

διπλωμάτης καριέρας/diplomate de carrière

δότης σπέρματος/donneur de sperme

εγκληματίας πολέμου/criminel de guerre

επόπτης γραμμών/arbitre de touche

ελεγκτής εναέριας κυκλοφορίας/contrôleur de la circulation aérienne

παιδί του δοκιμαστικού σωλήνα/bébé éprouvette

συγγενής πρώτου βαθμού/parent de premier degré

συνήγορος της πολιτικής αγωγής/avocat de la partie civile

Nhumcoll

απόσπασμα αγγαρείας/détachement de corvée

βουλή της αμφικτιονίας/Conseil des Amphictyons

γυμνάσιο θηλέων/école de filles

ελληνισμός της διασποράς/diaspora grecque

ορχήστρα δωματίου/orchestre de chambre

φυλακή ανηλίκων/prison pour mineurs

κέντρο ελευθέρων σπουδών/centre d'études libres

κυβέρνηση εθνικής ενότητας/gouvernement d'unité nationale

φυλακή υψίστης ασφαλείας/prison de haute sécurité

De manière générale, ces noms se réfèrent à des êtres humains afin de désigner leur fonction, leur activité professionnelle, leur statut, une qualité qui les caractérise etc. Les définitions lexicographiques que l'on retrouve dans les dictionnaires usuels sont du type : *ο άνθρωπος που V W/la personne qui V W* pour les Nhum et *το σύνολο των ανθρώπων που V W/l'ensemble de personnes qui V W* pour les Nhumcoll. Notons, cependant, que le plus souvent ce sont les noms simples uniquement qui reçoivent des définitions de ce type. Ceci est dû au fait que les noms composés ont rarement un statut lexicographique. Par conséquent, ils ne figurent presque jamais en tant qu'entrées indépendantes dans les dictionnaires usuels⁹⁹.

L'intuition sémantique est suffisante afin d'établir une liste des $N(E+DET:G)N:G$ et $NI(E+DET:G)A:GN2:G$ humains. Il s'agit d'un trait sémantique qui peut être considéré comme une propriété intrinsèque de ces noms.

Cependant, il y a certains noms qui ne désignent pas des êtres humains mais des êtres qui sont assimilés à ceux-ci. Il s'agit notamment des noms qui désignent des personnages imaginaires ou légendaires, des êtres surnaturels ou d'autres êtres spirituels (cf. B. Courtois 1994), comme par exemple : *θεός του πολέμου/dieu de la guerre*, *άγγελος Κυρίου/ange du Seigneur*, *τάγμα*

⁹⁹ Nous avons discuté ce problème dans II, 4.2.2.

αγγέλων/ordre des anges, *ιπότης της στρογγυλής τραπέζης*/chevalier de la table ronde. Toutefois, ces noms partagent avec les noms « humains » certaines propriétés formelles (cf. IV, 10.4.2).

10.4.2 Les critères formels

Les exemples précités montrent qu'il peut être assez difficile de décider de la classification sémantique d'un nom hors contexte et sur la base de l'intuition sémantique uniquement¹⁰⁰. Une autre méthode consiste donc à renforcer cette intuition sémantique par des tests formels, en examinant ce nom en situation dans une phrase. De cette façon, certaines propriétés formelles du nom peuvent être déduites et la classification sémantique du nom peut être effectuée en fonction de celles-ci. Plusieurs propriétés formelles peuvent être associées au trait humain. Tout d'abord, les noms « humains » peuvent être pronominalisés par *ποιος;/qui* ? dans les phrases interrogatives, par exemple :

- *Ποιος οργάνωσε αυτήν τη διαδήλωση;/Qui a organisé cette manifestation ?*
- *(Οι ανάπηροι πολέμου + ο ελληνισμός της διασποράς + το γυμνάσιο θηλέων + οι φυλακές ανηλίκων + *οι γραμμές των χεριών + *οι βάσεις δεδομένων)/(Les mutilés de guerre + la diaspora grecque + l'école de filles + les prisons pour mineurs + *les lignes de la main + *les bases de données)*
- *Ποιος προστατεύει τα παιδιά;/Qui protège les enfants ?*
- *(Άγγελοι Κυρίου + το προσωπικό ασφαλείας)/(Des anges du Seigneur + le personnel de sécurité)*

Ils partagent la distribution du pronom indéfini *κάποιος*/quelqu'un, par exemple :

*Αυτή η συμπεριφορά μπορεί να εξοργίσει (κάποιον + *κάτι)*

Ce comportement peut énerver (quelqu'un + *quelque chose)

*Αυτή η συμπεριφορά μπορεί να εξοργίσει (τον επόπτη γραμμών + τον ελεγκτή εναέριας κυκλοφορίας + το συνήγορο της πολιτικής αγωγής + τη φυλακή ανηλίκων + *τις γραμμές των χεριών + *τις βάσεις δεδομένων)*

¹⁰⁰ Pour la classification sémantique des *N (E+DET:G) N:G*, les classes d'objets, développées par G. Gross (1994), pourraient s'avérer très utiles. Sur les classes d'objets, cf. aussi D. Le Pesant ; M. Mathieu Colas (1998).

Ce comportement peut énerver (l'arbitre de touche + le contrôleur de circulation aérienne + l'avocat de la partie civile + la prison pour mineurs + *les lignes de la main + *les bases de données)

Cependant, ces propriétés formelles posent également un certain nombre de problèmes.

Tout d'abord, le pronom interrogatif *ποιος;/qui ?* ne sélectionne pas uniquement des noms humains ; il peut sélectionner également des noms animés (cf. IV, 10.5), étant donné que la sélection du nom est opérée par le verbe (ou le nom prédicatif). A titre d'illustration, citons les exemples suivants :

- *Ποιος πίνει νερό;/Qui est en train de boire de l'eau ?*
- *(Ο Νίκος + ο σκύλος)/(Nicolas + le chien)*
- *Ποιος έκανε αυτήν τη ζωγραφιά;/Qui a fait ce dessin ?*
- *(Ο Νίκος + *ο σκύλος)/(Nicolas + *le chien)*

Comme le signale A. Monceaux (1993 : 195), « les tests distributionnels ne sont pas des critères de classement suffisants en soi. Certains verbes de parole sélectionnent aussi bien des sujets humains que des noms de textes ou de support de texte, ou encore, jusqu'à une certaine limite, des noms d'animés non-humains, et l'acceptabilité de la question en *qui ?* est alors variable ». De plus, elle remarque qu'a priori on peut retrouver n'importe quel nom animé en position d'objet direct d'un verbe de sentiment, par exemple :

Η ζέστη εξουθένωσε (το Νίκο + την καμήλα)
La chaleur a exaspéré (Nicolas + le chameau)

On pourrait donc remarquer que la distinction entre les deux traits « animé » et « humain » en soi ne s'avère pas opératoire pour la description syntaxique, étant donné que les deux traits peuvent correspondre à la question formelle par *ποιος;/qui ?*.

En revanche, la distinction entre les deux traits se justifie dans la perspective d'un classement sémantique du lexique. Par exemple, des noms tels que *αρνάκι (E+του) γάλακτος*/agneau de lait, *γουρουνάκι (E+του) γάλακτος*/porc de lait etc. ou *γάτα Αγύρας*/chat angora, *γάτα Περσίας*/chat persan, *σκύλος Δαλματίας*/chien dalmatien, *σκύλος Αγίου Βερνάρδου*/chien Saint Bernard etc. ne désignent jamais des personnes. Pour cette raison, nous avons choisi, dans le cadre de cette étude, d'effectuer la distinction entre les deux traits animé et humain, bien que

les propriétés formelles présentées ci-dessus ne soient pas suffisantes pour marquer cette distinction de manière claire et précise.

Enfin, il faut noter que l'interprétation d'un nom donné dans une position syntaxique donnée est en partie déterminée par le contexte, ce qui constitue une autre limite des critères formels. Il s'agit d'emplois métaphoriques ou métonymiques dans lesquels des noms non-humains reçoivent la distribution habituelle des noms humains. Par exemple, dans des proverbes comme : *Είπε ο γάιδαρος τον πετεινό κεφάλι*/L'âne a appelé le coq idiot, *Η ομόνοια χτίζει σπίτι κι η διχόνοια το γκρεμίζει*/L'unité fait construire une maison et la discorde la détruit ou *Σηκώθηκαν τα πόδια να χτυπήσουν το κεφάλι*/Les pieds se sont levés pour battre la tête, les noms *ο γάιδαρος*/l'âne, *η ομόνοια*/l'unité, *η διχόνοια*/la discorde et *τα πόδια*/les pieds sont des noms *a priori* non-humains qui sont interprétés comme des humains.

Enfin, comme le signale A. Monceaux (1993 : 197), « il arrive que les emplois métonymiques ou métaphoriques de certains substantifs se lexicalisent et soient enregistrés dans les dictionnaires ». Nous en citons les exemples suivants que nous avons codé comme Nhum dans notre table du lexique-grammaire :

άχθος αρούρης

(« poids de terre » pour désigner une personne fainéante)

*κολόνα του σπιτιού*¹⁰¹

(« pilier de la maison » pour désigner la personne qui protège la famille)

ουρί του Παραδείσου

(« houri du paradis » pour désigner une très belle fille)

πεταλούδα της νύχτας

(« papillon de la nuit » pour désigner une femme de petite vertu)

Il s'agit d'un phénomène que l'on observe également dans le cadre des entités nommées. A titre indicatif, citons les exemples suivants :

γέρος του Μοριά

(« Vieux de Morée » pour désigner le héros national Théodore Kolokotronis)

θεά της σοφίας

(« déesse de la sagesse » pour désigner la déesse Athéna)

10.4.3 Les noms humains collectifs

Nous avons assigné le trait collectif uniquement aux noms humains. Rappelons que les propriétés formelles que nous avons décrites (cf. IV, 10.4.2) ne s'avèrent pas toujours opératoires en ce qui concerne la distinction entre individu et collectif :

— *Ποιος επέβαλε αυτά τα μέτρα;*/Qui a imposé ces mesures ?

— *Ο διευθυντής πωλήσεων*/Le directeur de ventes (individu)

— *Ο όμιλος επιχειρήσεων*/Le groupe d'entreprises (collectif)

Pour l'identification des Nhumcoll, nous nous sommes également servie d'un critère distributionnel : les noms collectifs peuvent être précédés de prédéterminants partitifs comme *ένα μέρος*/une partie de, *η πλειοψηφία*/la majorité de, Dnum % etc. (cf. M. Gross 1986, A. Monceaux 1993). A titre indicatif, citons l'exemple suivant :

Ένα μέρος (του ομίλου επιχειρήσεων + της ομάδας διάσωσης + του προσωπικού ασφαλείας + του συλλόγου καθηγητών) αναθεώρησε τις απόψεις του.

(Une partie (du groupe d'entreprises + de l'équipe de secours + du personnel de sécurité + de l'association des professeurs) est revenue sur ses positions.

L'application de ce critère facilite l'attribution du trait collectif tout en renforçant l'intuition sémantique sur ce point.

10.5 Les *N (E+DET:G) N:G* et les *NI (E+DET:G) A:G N2:G* animés non-humains

Nous avons montré dans IV, 10.4.2 que les noms animés ne se distinguent pas formellement des noms humains, car ils ne correspondent pas à des propriétés syntaxiques particulières. Les *N (E+DET:G) N:G* et *NI (E+DET:G) A:G N2:G* animés que nous avons recensés constituent, dans la plupart des cas, des dénominations de races ou espèces d'animaux. De manière

¹⁰¹ Précisons qu'une séquence comme *κολόνα του σπιτιού*/pilier de la maison peut également désigner un objet concret. Cependant, dans ce cas, il s'agit d'une séquence libre, qui ne figure pas dans notre dictionnaire et nos tables du lexique-grammaire.

générale, ils ne posent pas de réels problèmes de prise de décision. A titre indicatif, citons les exemples suivants :

(αίγα + γάτα + κουνέλι) Αγκύρας

((chèvre + chat + lapin) angora)

άλογο κούρσας

(cheval de course)

ζώο του δάσους

(animal sauvage)

σκύλος σαλονιού

(chien de salon)

ψάρι του γλυκού νερού

(poisson d'eau douce)

σκύλος του Αγίου Βερνάρδου

(chien Saint Bernard)

Le phénomène des emplois métaphoriques ou métonymiques (cf. IV, 10.4.2) s'observe également au sein des noms animés. A titre d'illustration, citons les exemples suivants :

πλοίο της ερήμου

(« bateau du désert » pour désigner le chameau)

σταυρός της θάλασσας

(« croix de la mer » pour désigner l'étoile de mer)

Notons, enfin, que nous n'avons pas codé le trait animal collectif.

10.6 Les *N (E+DET:G) N:G* et les *NI (E+DET:G) A:G N2:G* concrets

10.6.1 La notion sémantique « concret »

La notion sémantique de nom « concret » correspond à un nom, qui désigne un objet du monde sensible, soit un être soit un objet dont l'existence est appréhendée par des sens tels que le toucher ou la vue (cf. A. Monceaux 1993 : 205).

Rappelons que les traits utilisés sont des traits exclusifs. Ceci signifie que les noms composés marqués « humain » ou « animé non-humain » ont été exclus de la classe des « concrets ». Comme le note A. Monceaux 1993 : 205, « cette remarque implique qu'une typologie des noms humains et animés sera sans doute nécessaire pour prévoir leur possibilité d'apparition dans une position d'argument notée *Nconc* dans le lexique grammairal ». A titre indicatif, citons l'exemple suivant :

Έφαγαν τον αρχηγό τους

Ils ont mangé leur chef

où le nom *αρχηγός*/chef est employé dans un contexte qui a trait au corps de l'être en question. On peut considérer qu'il y a eu, dans ce cas, réduction d'un groupe nominal complexe comportant un nom approprié de partie du corps tel que *σώμα*/corps, *σάρκα*/chair etc. :

Έφαγαν (το σώμα + τη σάρκα) του αρχηγού τους

Ils ont mangé (le corps + la chair) de leur chef

« Du fait de l'extension donnée au trait « humain », l'inclusion généralisée de la classe des humains à celle des concrets serait fautive, la classe *Nhum* comprenant des noms tels que personne morale, qui ne réfère jamais à la partie matérielle d'un être humain » (A. Monceaux 1993 : 205).

Nous n'avons pas pu identifier des propriétés formelles qui pourraient être associées à la notion du nom concret. En effet l'interprétation d'un substantif dépend de son contexte. De manière générale, les noms composés désignant des objets concrets pourraient toujours recevoir une interprétation générique abstraite, par exemple :

Ο ανιχνευτής ψεύδους είναι μία καταπληκτική εφεύρεση

Le détecteur de mensonges est une invention formidable

Το γάλα κονσέρβας δεν είναι θρεπτικό για τον οργανισμό

Le lait en boîte n'est pas nutritif pour l'organisme

« Les conditions de cette interprétation générique apparaissent comme très complexes. L'analyse de certaines de ces phrases pourrait mettre en jeu des enchâssements et réductions de propositions subordonnées : le nom concret apparaît dans une position d'argument non restreint » (A. Monceaux 1993 : 205).

Ces analyses devraient être formalisées. Toutefois, elles ne rendent pas compte de toutes les possibilités d'interprétation d'un nom concret.

Dans le cadre de cette étude, le trait « concret » a été assigné à tous les *N (E+DET:G) N:G* et *NI (E+DET:G) A:G N2:G* qui, pris isolément, peuvent référer à un objet du monde sensible, à savoir un objet dont l'existence peut être appréhendée par des sens tels que le toucher ou la vue, à l'exception des êtres humains et animés (cf. A. Monceaux 1993 : 206). A titre indicatif, citons quelques exemples de *N (E+DET:G) N:G* et *NI (E+DET:G) A:G N2:G* codés comme « concrets » :

ανιχνευτής ψεύδους

(détecteur de mensonges)

γάλα κονσέρβας

(lait en boîte)

γραμμή της ζωής

(ligne de vie)

γυαλιά ηλίου

(lunettes de soleil)

δοχείο νυκτός

(pot de chambre)

ζώνη ασφαλείας

(ceinture de sécurité)

αλεύρι ολικής άλεσης
(farine complète)

γάλα μακράς διαρκείας
(lait longue durée)

μάτι της ηλεκτρικής κουζίνας
(chauffe-plats)

Notons que les noms de partie du corps, comme par exemple :

ημισφαίριο του εγκεφάλου
(hémisphère cérébral)

κόλπος της καρδιάς
(oreillette)

τράχηλος της μήτρας
(col de l'utérus)

ont également été codés comme « concrets ». En revanche, les noms désignant une maladie, comme par exemple :

θρόμβωση της καρωτίδας
(thrombose de la carotide)

ανεύρυσμα αορτής
(anévrisme de l'aorte)

διάσειση εγκεφάλου
(commotion cérébrale)

font partie de la classe résiduelle des « noms abstraits ». Dans un premier temps, ils pourraient être étudiés avec les noms prédicatifs avec verbe support *έχω/avoir*.

Enfin, en ce qui concerne le codage des noms de texte et/ou de support de texte, nous les avons marqués de deux traits à chaque fois : celui de nom « concret », puisqu'ils peuvent désigner un objet concret dont l'existence peut être appréhendée par le toucher ou la vue, mais aussi celui de la classe résiduelle des « abstraits », étant donné que, dans un autre contexte, ils

peuvent également avoir un emploi prédicatif avec le verbe support *κάνω*/faire ou ses variantes lexicales *ετοιμάζω*/préparer, *συντάσσω*/rédiger, *γράφω*/écrire. A titre d'illustration, citons quelques exemples de $N (E+DET:G) N:G$ de ce type :

μελέτη σκοπιμότητας

(étude de faisabilité)

προσχέδιο νόμου

(projet de loi)

σύμβαση εργασίας

(contrat de travail)

πρόταση μομφής

(motion de censure)

10.7 La classe des noms abstraits

Dans la pratique, l'opposition concret/abstrait se heurte à de nombreuses difficultés d'application, un même mot pouvant être abstrait ou concret selon son contexte. Le critère « concret » n'est souvent pas opératoire dans la mesure où il dépend de l'interprétation que l'on donne au mot *concret* lui-même. [...] La multiplicité de sens que peut prendre un mot en dehors de tout contexte est la source des difficultés du marquage » (B. Courtois 1994).

Les $N (E+DET:G) N:G$ qui ne sont pas marqués des traits sémantiques « humain », « animé » et « concret » constituent une classe résiduelle (cf. A. Monceaux 1993 : 210). Nous les considérons comme des noms abstraits. Une grande partie de ces noms sont des noms prédicatifs rattachés à différents verbes supports (*κάνω*/faire, *έχω*/avoir etc.). Les noms « abstraits » de la classe résiduelle ne constituent pas une classe homogène. A titre d'illustration, citons quelques exemples de $N (E+DET:G) N:G$ abstraits :

αθανασία της ψυχής

(immortalité de l'âme)

απεργία πείνας

(grève de la faim)

απόπειρα δολοφονίας

(tentative d'assassinat)

διαχείριση κρίσης

(gestion de crise)

σύμπλεγμα κατωτερότητας

(complexe d'infériorité)

μανία καταδίωξης

(manie de persécution)

Une étude plus détaillée des noms « abstraits » reste donc à faire.

Comme nous l'avons vu, les traits sémantiques servent à séparer les différents sens de termes polysémiques. Cependant, le travail de marquage sémantique fait apparaître la difficulté de définition des traits ainsi que la difficulté d'assigner un trait à chaque entrée du dictionnaire ou de la table du lexique-grammaire. Cette difficulté est due au fait que chaque entrée est considérée en dehors de tout contexte. En théorie, un lexique-grammaire des noms et de leurs emplois permettrait de lever cette difficulté, car alors le marquage des noms pourrait être systématisé (B. Courtois 1994).

CINQUIEME PARTIE : TRAITEMENT AUTOMATIQUE

Chapitre 11. L'analyse automatique des textes grecs

11.1 Introduction

Concernant la reconnaissance automatique des expressions multi-mots, les recherches actuelles en TAL suivent trois grandes approches :

- L'approche linguistique ou symbolique, qui s'appuie sur la description et la modélisation explicites et exhaustives de la langue naturelle. Cette méthode consiste à utiliser une base de règles construites à la main, souvent exprimées sous forme de graphes d'automates finis (cf. D. Perrin 1989, S.-C. Kleen 1956) et exploite de manière précise et exhaustive le contexte local dans lequel apparaissent les expressions multi-mots. Dans le cadre de cette approche, la construction de ressources linguistiques est une étape d'une importance capitale. L'avantage de cette approche est qu'elle est adaptée au profil des linguistes qui travaillent dans ce domaine. Du point de vue applicatif, un système qui repose sur l'approche symbolique atteint une précision importante (E. Roche et Y. Schabes 1997, J. Hobbs et al. 1996). De plus, de tels systèmes se caractérisent par une lisibilité facile. En revanche, l'impossibilité d'atteindre une couverture lexicale exhaustive constitue leur inconvénient majeur. A ce problème s'ajoute aussi la difficulté importante au niveau de la gestion et de la maintenance des ressources. Toutefois, une fois ces problèmes identifiés, des solutions peuvent être conçues pour faciliter la tâche de la mise à jour. GraalWeb est un système de bibliothèque décentralisée qui permet de partager et gérer en-ligne des grammaires locales de descriptions linguistiques (M. Constant 2007). De plus, des outils d'enrichissement et de validation automatisés des dictionnaires électroniques morphologiques ont été développés pour le grec moderne (M. Constant et A. Yannacopoulou 2003).
- L'approche statistique, qui n'utilise que peu ou pas d'informations linguistiques explicites. Elle consiste en l'acquisition de connaissances par apprentissage automatique à partir de corpus annotés manuellement par des experts. Le système est ensuite entraîné de manière supervisée ou non-supervisée à analyser des textes. La désambiguïsation des unités lexicales s'opère en déterminant la plus probable des interprétations possibles. Le principal inconvénient de cette approche est le coût

d'annotation, vu qu'il y a très peu de corpus de référence disponibles et de taille conséquente. En ce qui concerne les systèmes basés sur une approche statistique, ils ont du mal à reconnaître des données qui sont rares et dispersées. En revanche, leur grand avantage est la robustesse : le système fournit toujours une réponse puisqu'il repose sur une méthode fondée sur les fréquences, c'est-à-dire sur l'étude des cooccurrences d'objets linguistiques (n-gram).

- Enfin, l'approche hybride, qui combine l'approche symbolique et l'approche statistique. Elle couple le traitement statistique avec un minimum de ressources linguistiques afin de traiter des phénomènes qui présentent une fréquence restreinte dans le but d'effectuer l'entraînement à l'apprentissage automatique. Pour ce qui est du français, une ressource importante issue de méthodes hybrides est le corpus arboré *FRENCH TREEBANK* (A. Abeillé et al. 2003) qui est annoté à trois niveaux : morphosyntaxique (tagging), syntagmatique (chunking) et fonctionnel (parsing). L'annotation morphosyntaxique a été effectuée à l'aide des dictionnaires électroniques morphologiques. Les mots composés ont été annotés automatiquement dans le corpus à l'aide d'un programme basé sur des expressions régulières pour reconnaître les nombres écrits en lettres (e.g. mille deux cent quatre vingt-seize) et d'un dictionnaire pour identifier les autres types de mots composés (incluant notamment des extraits des dictionnaires accessibles sous INTEX). Citons aussi les travaux de B. Daille (1995, 1996), qui utilise des méthodes combinées (linguistiques et statistiques) pour l'extraction de terminologie.

Rappelons que, dans le cadre de la présente étude, nous avons effectué le recensement (cf. II, 4.2.1-4.2.3), l'étude de la structure lexicale interne (cf. III, 6 et 7), l'étude syntaxico-sémantique (cf. IV, 9-10) et la représentation formalisée (cf. V, 11.3.1-11.3.3) des $N(E+DET:G)N:G$ du grec moderne en vue de leur reconnaissance automatique dans les textes grecs. La méthode que nous appliquons ici repose sur l'approche linguistique et n'utilise aucune information statistique. Cette méthode nécessite la construction de ressources linguistiques de large couverture, représentées sous un formalisme aisément traitable.

Dans cette partie, nous abordons d'abord les problèmes de reconnaissance lexicale automatique propres aux $N(E+DET:G)N:G$ grecs. Ensuite, nous présentons les méthodes de représentation formalisée que nous avons choisies pour le traitement automatique de chacune des sous-catégories identifiées. Nous illustrons nos choix par des exemples de nos

dictionnaires électroniques morphologiques, nos grammaires locales et nos tables du lexique-grammaire. Enfin, nous testons nos outils à l'aide d'Unitex et nous commentons les résultats obtenus.

11.2 Reconnaissance lexicale des $N(E+DET:G)N:G$ dans les textes

Comme le signale S. Voyatzi (2006 : 351), depuis longtemps, on insiste sur la nécessité d'identifier les différents types d'expressions figées¹⁰² dans le cadre de différentes tâches du TAL, mais leur nombre élevé et leur caractère mouvant rendent relativement complexe une telle entreprise. La majorité des systèmes existants ne les traitent pas de façon satisfaisante, « car on manque de méthodes générales pour les reconnaître » (E. Laporte 1988 : 117).

Par « reconnaissance lexicale » des $N(E+DET:G)N:G$ grecs, nous entendons leur identification dans les textes, indépendamment de leur position syntaxique.

Les noms composés sont, par définition, des formes multi-mots. Dans la majorité des cas, pour leur identification dans les textes, des bases de données qui fournissent leur forme *in extenso* et toutes leurs propriétés sont nécessaires. Cependant, la reconnaissance lexicale automatique des noms composés est souvent compliquée. Un des principaux problèmes qui gênent la reconnaissance lexicale automatique des $N(E+DET:G)N:G$ est de nature linguistique et est propre à leur structure lexicale interne : la présence de différents types de variantes¹⁰³, notamment, les variantes lexicales, orthographiques, syntaxiques etc.

11.2.1 Problèmes de variantes des $N(E+DET:G)N:G$

Nous avons déjà montré qu'un nombre considérable de $N(E+DET:G)N:G$ est susceptible de variation lexicale, orthographique, phonologique, syntaxique ou substitution par des formes abrégées. Nous avons étudié en détail ces phénomènes de variation dans II, 5. Rappelons ici quelques cas intéressants :

- variantes flexionnelles :

αντιρρησίας συνειδήσεως (variante savante)

¹⁰² Le terme d'« expression figée » est utilisé ici de manière à désigner toute unité lexicale qui est traitée dans le cadre du figement (*i.e.* noms composés, proverbes, adverbies figés, etc.).

¹⁰³ Sur les variantes des $N(E+DET:G)N:G$ grecs, cf. II, 5.

αντιρρησίας συνείδησης (variante démotique)
(objecteur de conscience)

- variantes orthographiques :

(*δικλίδα + δικλείδα*) ασφαλείας
(soupape de sûreté)

- substitution par des formes abrégées :

τεστ Παπανικολάου (forme complète)
τεστ παπ (forme abrégée)
(pap test) (domaine : Médecine)

- variantes phonologiques :

(*λωρίδα + λουρίδα*) κυκλοφορίας
(voie de circulation)

- variantes syntaxiques :

$N \quad N:G \quad = \quad N-a \quad N$
πύελος νεφρού = νεφρική πύελος
(bassinets rénal)

Comme nous l'avons déjà signalé tout au long de II, 5, la méthode que nous proposons pour la représentation formalisée des variantes lexicales et orthographiques est le dictionnaire électronique morphologique, que nous présenterons plus explicitement par la suite (cf. V, 11.3.1). A titre indicatif, citons l'exemple suivant¹⁰⁴ :

δικλείδα ασφαλείας, .N+NNG:Nfs:Afs:Vfs+Nconc
δικλίδα ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Nfs:Afs:Vfs+Nconc
δικλείδας ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Gfs+Nconc
δικλίδας ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Gfs+Nconc
δικλείδες ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Nfp:Afp:Vfp+Nconc
δικλίδες ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Nfp:Afp:Vfp+Nconc
δικλείδων ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Gfp+Nconc
δικλίδων ασφαλείας, *δικλείδα ασφαλείας*.N+NNG:Gfp+Nconc

¹⁰⁴ Pour l'instant, nous avons effectué manuellement la mise en relation des variantes orthographiques de cet exemple. L'élaboration du programme pour le traitement automatique des variantes orthographiques dans le dictionnaire électronique morphologique est en cours.

Dans l'exemple précité, les deux variantes orthographiques :

(δικλίδα + δικλείδα) ασφαλείας

(soupape de sûreté)

sont reliées par la même forme canonique, en l'occurrence δικλείδα ασφαλείας/soupape de sûreté, dans le dictionnaire électronique morphologique.

En ce qui concerne les variantes syntaxiques, nous les avons traitées à l'aide des tables du lexique-grammaire¹⁰⁵, que nous avons par la suite converties automatiquement en graphes de transducteurs finis. A titre indicatif, citons deux exemples tirés de notre table du lexique-grammaire :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S

Figure 3 : Exemple de représentation formalisée pour les variantes de type AN

Dans cet exemple, le $N(E+DET:G) N:G$ διάσειση του εγκεφάλου/commotion cérébrale a une variante de type AN, à savoir εγκεφαλική διάσειση/commotion cérébrale. Ce type de variante syntaxique est décrit dans la colonne I, intitulée $N DET:G N:G=N:G-a N$. Nous avons donc marqué « + » pour cette entrée dans la colonne I. Ensuite, dans la colonne J, nous avons noté l'adjectif morphologiquement lié au $N:G$ qui donne lieu à la formation du AN.

Précisons que, à l'heure actuelle, la mise en relation des $N(E+DET:G) N:G$ avec les variantes de type AN est effectuée uniquement dans les tables du lexique-grammaire. Cette mise en relation devrait être effectuée dans le dictionnaire électronique morphologique aussi.

Examinons maintenant l'exemple suivant :

¹⁰⁵ Les tables du lexique-grammaire sont présentées plus explicitement dans V, 11.3.3.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S		
N	DET:G	N:G	N DET:G N:G	N N:G	DET:G=E	DET:G N:G	Namb	N DET:G N:G=N:G-a N		N:G-a	N DET:G N:G=Ns		Ns	N DET:G N:G=Npréd	N DET:G N:G=Nhum	N DET:G N:G=Nhumcoll	N DET:G N:G=Hanim	N DET:G N:G=Nconc	N DET:G N:G=Nabs	N DET:G N:G=TS
ζώνη	του	ευρώ	+	-	+	-	-	-	ἅ	+	ευρωζώνη	-	-	-	+	-	-	-		

Figure 4 : Exemple de représentation formalisée pour les variantes soudées

Le $N(E+DET:G)N:G$ ζώνη του ευρώ/zone euro présente une variante soudée, à savoir *ευρωζώνη/eurozone*. Ce type de variante syntaxique est décrit dans la colonne K, intitulée $N DET:G N:G=N:G$. Nous avons donc marqué « + » pour cette entrée dans la colonne K. Ensuite, dans la colonne L, nous avons noté le nom soudé (Ns) correspondant, qui constitue la forme variée du $N(E+DET:G)N:G$.

Tout comme pour les variantes de type AN , à l'heure actuelle, la mise en relation des $N(E+DET:G)N:G$ avec les variantes soudées est effectuée uniquement dans les tables du lexique-grammaire. Cependant, notre but est d'effectuer cette mise en relation dans le dictionnaire électronique morphologique aussi.

11.3 Méthodes de reconnaissance lexicale automatique des $N(E+DET:G)N:G$

De manière générale, les systèmes symboliques utilisent deux méthodes pour la reconnaissance lexicale automatique des unités lexicales :

- des dictionnaires électroniques morphologiques ;
- des règles de grammaire exprimées sous forme de graphes d'automates à états finis récursifs (RTN), d'expressions régulières ou de règles construites *ad hoc*.

Dans le cadre de cette étude, pour la reconnaissance lexicale automatique des $N(E+DET:G)N:G$ nous avons élaboré les ressources suivantes : des dictionnaires électroniques morphologiques, des grammaires locales et des tables du lexique-grammaire, que nous avons par la suite converties automatiquement en graphes d'automates à états finis à l'aide d'Unitex.

11.3.1 Les dictionnaires électroniques morphologiques

Rappelons que cette méthode est couramment utilisée pour la reconnaissance lexicale automatique des unités lexicales simples et des unités lexicales multi-mots. Il s'agit d'une méthode de reconnaissance particulièrement adéquate pour les mots ou les suites de mots susceptibles de variations lexicales ou orthographiques. C'est donc la méthode que nous avons choisie pour la reconnaissance des noms composés entièrement figés, les termes de spécialité, les noms composés qui font partie d'une phrase simple figée et qui peuvent avoir un emploi autonome en dehors de la phrase figée ainsi que pour les syntagmes nominaux semi-figés (ou productifs) dont le paradigme est relativement restreint (cf. III, 6.4, 6.7 et 6.8).

Les dictionnaires électroniques morphologiques que nous avons élaborés décrivent la morphologie, la structure lexicale et les variantes lexicales (flexionnelles), orthographiques et phonologiques¹⁰⁶ des $N(E+DET:G)N:G$ et des $NI(E+DET:G)A:GN2:G$. Ils font partie du dictionnaire électronique morphologique des mots composés du grec moderne (cf. T. Kyriacopoulou *et al.* 2002), qui est analogue au système DELA¹⁰⁷ français (cf. B. Courtois 1990) et destiné à l'analyse automatique des textes écrits.

L'organisation d'un dictionnaire électronique morphologique, visant à la reconnaissance lexicale automatique, repose tout d'abord sur le concept de mot, en tant qu'unité élémentaire de traitement. Rappelons qu'en ce qui concerne le grec moderne, nous avons adopté la définition formelle proposée par M. Silberztein (1990)¹⁰⁸.

En nous basant, d'une part, sur cette distinction formelle et, d'autre part, sur notre liste de $N(E+DET:G)N:G$ (cf. II, 4.2.1-4.2.3), nous avons élaboré deux dictionnaires électroniques¹⁰⁹, à savoir :

¹⁰⁶ Les variantes syntaxiques sont traités à l'aide des tables du lexique-grammaire, cf. V, 11.3.3.

¹⁰⁷ Dictionnaire Electronique du LADL.

¹⁰⁸ Cf. aussi G. Gross (1996a).

¹⁰⁹ Notons que les deux dictionnaires seront intégrés au dictionnaire électronique morphologique des noms composés. Nous avons choisi d'élaborer deux dictionnaires pour mieux étudier chacune des catégories concernées.

- le dictionnaire électronique morphologique des $N (E+DET:G) N:G$ binaires ;
- le dictionnaire électronique morphologique des $N (E+DET:G) N:G$ composés complexes ou surcomposés (rappelons que nous n'avons élaboré que le dictionnaire des noms composés complexes ou surcomposés de structure $NI (E+DET:G) A:G N2:G$).

Le dictionnaire électronique des unités lexicales multi-mots doit contenir toutes les informations nécessaires à l'automatisation de la génération de formes fléchies. Les noms composés doivent être décrits morphologiquement du point de vue de leur flexion. Pour la flexion des noms composés, nous nous sommes basée sur les mots simples et nous avons appliqué des filtres pour éliminer les formes non acceptables ou agrammaticales. Ainsi, pour la flexion des $N (E+DET:G) N:G$, nous avons pris en compte toutes les caractéristiques morphologiques de la langue grecque, à savoir les trois genres (masculin, féminin, neutre), les deux nombres (singulier, pluriel) et les quatre cas (nominatif, génitif, accusatif, vocatif). Nous avons également pris en compte le déplacement obligatoire de l'accent qui s'observe lors de la flexion de certains noms (cf. T. Kyriacopoulou 2003 : 77-79).

11.3.1.1 Structure des entrées

Le DELAC¹¹⁰ est une base de données morphologiques. Pour les entrées du DELAC des $N (E+DET:G) N:G$ et des $NI (E+DET:G) A:G N2:G$, la structure de chaque entrée consiste en :

- un code alphanumérique (e.g. N301), qui contient un symbole de partie du discours (en l'occurrence N qui désigne un nom) et une classe flexionnelle, pour chaque composant des unités lexicales multi-mots ;
- le code N+NNG¹¹¹, qui spécifie que l'entrée dans son ensemble est un nom, et plus particulièrement un nom composé défini par la structure morphosyntaxique $N (E+DET:G) N:G$;
- un filtre de restriction de formes.

¹¹⁰ Dictionnaire électronique morphologique des formes canoniques des noms composés.

¹¹¹ Notons que l'utilisation du code « NNG » pourrait s'avérer utile ultérieurement lors de l'élaboration de filtres plus spécifiques ou d'un codage automatique.

Plus explicitement, pour la flexion des noms composés deux types de filtres sont utilisés : des filtres sur le genre et des filtres sur la déclinaison. Dans les deux cas, il faut distinguer les filtres génériques, qui s'appliquent à l'ensemble des composants d'un mot composé, et les filtres spécifiques, qui s'appliquent à un composant particulier repéré par sa position dans le mot composé. Les filtres génériques sont M, F et N et signifient respectivement que le composé est toujours au masculin, féminin et neutre.

Les filtres sur la déclinaison peuvent se diviser en trois catégories :

- Les filtres sur le nombre : ils indiquent si le composant (ou le nom composé dans son ensemble) n'existe qu'au singulier ou qu'au pluriel. Ainsi, pour les filtres génériques, on aura S pour le singulier et P pour le pluriel. On retrouve la notation S_i et P_i pour le $i^{\text{ème}}$ composant.
- Les filtres sur les cas : ils indiquent si le nominatif, le génitif, l'accusatif et le vocatif sont autorisés ou non. N, G, A, V signifient respectivement au nominatif, génitif, accusatif, vocatif singulier et pluriel. Les filtres spécifiques fonctionnent avec la notation $-N_i$, $-G_i$, $-A_i$, $-V_i$.
- Les filtres peuvent se combiner. Par exemple, pour un $N(E+DET:G)N:G$, les codes S,-GS2 signifient que les deux mots simples sont toujours au singulier et que le deuxième est forcément au génitif singulier. Il y aura donc 4 formes différentes : -NS1 -GS2, -GS1 -GS2, -AS1 -GS2 et -VS1 -GS2 (cf. T. Kyriacopoulou 2003: 88).

Dans le dictionnaire électronique morphologique que nous avons élaboré, nous avons ajouté manuellement le trait syntaxico-sémantique *Npréd*, lorsque il s'agit d'un nom prédictif, et les traits sémantiques suivants¹¹² :

- *TS*, quand il s'agit d'un terme de spécialité ;
- *Nhum*, pour désigner les noms humains ;
- *Nhumcoll*, pour désigner les noms humains collectifs ;
- *Nanim*, pour désigner les noms animés (non-humains) ;

¹¹² L'élaboration du programme pour l'ajout automatique de ces traits est en cours.

- *Nconc*, pour désigner les noms concrets ;
- *Nabs*, pour désigner les noms abstraits.

Les exemples ci-dessous illustrent la structure des entrées du DELAC :

αεροπλάνο.N311 μάχης.N253,N+NNG,-GS2+Nconc
απαλοϊφή.N251 παρονομαστών.N190,N+NNG,S,-GP2+Nabs+TS+Npréd

Dans les exemples précités, N311, N253, N251 et N190 indiquent des noms respectivement de classe flexionnelle 311, 253, 251 et 190 ; -GS2 désigne que le deuxième constituant de cette entrée se met obligatoirement au génitif singulier, -GP2 désigne que le deuxième constituant de cette entrée se met obligatoirement au génitif pluriel ; S est un autre filtre de restriction qui désigne que cette entrée n'a pas de pluriel ; Nconc et Nabs sont des traits sémantiques élémentaires qui indiquent que l'entrée est respectivement un nom concret et un nom abstrait ; enfin, TS et Npréd s'utilisent respectivement pour le terme de spécialité et le nom prédicatif. Sur ces exemples, on voit que les entrées figurent sous leur forme canonique, à savoir au nominatif singulier.

11.3.1.2 Génération automatique des formes fléchies

Le DELAC permet d'obtenir le DELACF, qui constitue le dictionnaire électronique morphologique des formes fléchies, à l'aide de GenereFlexion, le programme de flexion élaboré par S. Mrabti (T. Kyriacopoulou *et al.* 2002). Le programme de flexion a été présenté par T. Kyriacopoulou *et al.* (2002), il n'est donc pas nécessaire d'y revenir ici en détail. Nous nous contentons de rappeler que pour la génération automatique de toutes les formes fléchies le programme de flexion utilise deux fichiers, à savoir : le fichier des entrées (cf. V, 11.3.1.1) des *N (E+DET:G) N:G* (DELAC) et le fichier des codes flexionnels (cf. V, 11.3.1.3.1). Un troisième fichier, le fichier des résultats (DELACF), est produit automatiquement par le programme de flexion et contient toutes les formes fléchies. Ainsi, après avoir traité l'extrait du DELAC présenté dans V, 11.3.1.1, GenereFlexion produit le DELACF pour les entrées concernées :

αεροπλάνο μάχης,.N+NNG:Nns:Ans:Vns+Nconc
αεροπλάνου μάχης,αεροπλάνο μάχης.N+NNG:Gns+Nconc
αεροπλάνα μάχης,αεροπλάνο μάχης.N+NNG:Nnp:Anp:Vnp+Nconc
αεροπλάνων μάχης,αεροπλάνο μάχης.N+NNG:Gnp+Nconc
απαλοϊφή παρονομαστών,.N+NNG:Nfs:Afs:Vfs+Nabs+TS+Npréd
απαλοϊφής παρονομαστών,απαλοϊφή παρονομαστών.N+NNG:Gfs+Nabs+TS+Npréd

On voit que chaque entrée de notre dictionnaire morphologique comprend la forme fléchie du nom composé (forme précédant la virgule), accompagnée de sa forme canonique (forme suivant la virgule)¹¹³ et des informations morphologiques concernant le cas, le genre et le nombre. Nous explicitons ces codes ci-après :

Nms ; Nmp	nominatif masculin singulier ; nominatif masculin pluriel
Nfs ; Nfp	nominatif féminin singulier ; nominatif féminin pluriel
Nns ; Nnp	nominatif neutre singulier ; nominatif neutre pluriel
Gms ; Gmp	génitif masculin singulier ; génitif masculin pluriel
Gfs ; Gfp	génitif féminin singulier ; génitif féminin pluriel
Gns ; Gnp	génitif neutre singulier ; génitif neutre pluriel
Ams ; Amp	accusatif masculin singulier ; accusatif masculin pluriel
Afs ; Afp	accusatif féminin singulier ; accusatif féminin pluriel
Ans ; Anp	accusatif neutre singulier ; accusatif neutre pluriel
Vms ; Vmp	vocatif masculin singulier ; vocatif masculin pluriel
Vfs ; Vfp	vocatif féminin singulier ; vocatif féminin pluriel
Vns ; Vnp	vocatif neutre singulier ; vocatif neutre pluriel

11.3.1.3 La flexion des unités lexicales multi-mots

La flexion des unités lexicales multi-mots est basée, d'une part, sur les codes flexionnels de leurs composants et, d'autre part, sur les filtres de restriction pour éliminer les formes non acceptables (cf. aussi 11.3.1.1). Dans cette section, nous présentons les particularités de la flexion des *N (E+DET:G) N:G*.

¹¹³ Dans le cas où une entrée du DELACF ne présente qu'une seule forme, e.g.: αντιρρησίας συνείδησης, .N+NNG:Nms, Nhum, cette forme est identique à la forme canonique et donc elle n'est pas répétée.

11.3.1.3.1 Déplacement de l'accent

Comme nous l'avons déjà signalé, pour la flexion des $N (E+DET:G) N:G$, nous nous sommes basée sur les classes flexionnelles de leurs composants en utilisant des filtres de restriction. Les codes flexionnels regroupent des mots qui ont des caractéristiques morphologiques communes.

Dans certains cas, l'accent se déplace lors de la flexion. Par exemple, pour le mot simple *δικαίωμα*/droit, nous avons:

Entrée du dictionnaire des mots simples :

δικαίω"μα.N365,N

Fichier Codes Flexionnels :

N365.3.α, "'''°ατος,α,α, "'''°ατα, °άτων, "'''°ατα, "'''°ατα

Nous observons donc qu'au nominatif pluriel un déplacement et une suppression de l'accent initial ont lieu afin d'obtenir *δικαιώματα*/droits.

Le mot *δικαίωμα*/droit utilisé dans le $N (E+DET:G) N:G$ *δικαιώματα του ανθρώπου*/droits de l'homme est un cas particulier. Il est employé au nominatif pluriel et l'accent se trouve déjà sur la lettre *ω* (*δικαίω"μα*) qui doit porter l'accent. Or, dans la classe flexionnelle, il est indiqué qu'au nominatif pluriel il y a un déplacement et une suppression de l'accent (*"'''°ατα*). En faisant une entrée :

δικαιώματα.N365 του.DET ανθρώ"που.N125,N+NNG,P

le programme ne trouve pas d'accent à déplacer mais il supprime l'accent existant. Pour résoudre ce type de problèmes, nous sommes obligée :

a) d'indiquer, dans le DELAC, la voyelle qui va porter l'accent même s'il s'agit de la lettre qui porte déjà l'accent :

δικαιώ"ματα.N365 του.DET ανθρώ"που.N125,N+NNG,P

b) d'enlever l'accent avant d'en ajouter un nouveau. Si les deux opérations ne sont pas faites dans cet ordre, le mot se retrouvera sans accent (cf. T. Kyriacopoulou 2003 : 90).

11.3.1.3.2 *Genre des N (E+DET:G) N:G*

La règle prévoit que le genre est donné par le premier nom du nom composé si celui-ci contient un nom (cf. T. Kyriacopoulou 2003 : 91). Dans le cas des *N (E+DET:G) N:G* la règle s'applique parfaitement, car le genre du *N (E+DET:G) N:G* est hérité du genre du nom-tête.

11.3.1.3.3 *Les variantes*

Un nombre important de variantes lexicales (flexionnelles), orthographiques et phonologiques (cf. T. Kyriacopoulou 2003: 77-79) figurent dans le dictionnaire électronique morphologique. Elles sont reliées par la même « forme canonique » (cf. T. Kyriacopoulou 1990, T. Kyriacopoulou *et al.* 2002). La notion de « forme canonique » attestée relève du niveau morphologique des langues et désigne, en pratique, la forme d'une unité lexicale telle qu'elle figure en entrée dans les ouvrages lexicographiques. Selon cet usage, la « forme canonique » des *N (E+DET:G) N:G* du grec moderne est la forme au nominatif du singulier, *i.e.* *ανιχνευτής ψεύδους*/détecteur de mensonges.

11.3.1.4 **Application des dictionnaires et résultats obtenus**

Le DELAC des *N (E+DET:G) N:G* contient à ce jour 2 460 entrées, qui correspondent à 11 195 formes fléchies. Le DELAC des *N1 (E+DET:G) A:G N2:G* contient actuellement 244 entrées, qui correspondent à 1 346 formes fléchies. Nous présentons des extraits représentatifs de nos dictionnaires en Annexes 1 et 2.

Nous présentons ci-après des extraits des résultats que nous avons obtenus à l'aide d'Unitex après l'application de nos dictionnaires électroniques morphologiques dans un corpus (journal TA NEA, janvier 2003).

i) Concordances du DELACF des *N (E+DET:G) N:G* :

ης αγγίζει, τέλος, λεπτά ζητήματα του [Κράτους Δικαίου](#).{S} Μήπως στον βωμό της δημ
ς της χώρας τους, αφού απελάθηκαν. {S}[Ομάδες διάσωσης](#) εξακολούθησαν και χθες να αναζητούν τον
στο Αφγανιστάν. «Εξυπνες» βόμβες, νέα [άρματα μάχης](#) και νέα ελικόπτερα.{S} Ο πόλεμος που προετ
με» του 23χρονου ράπερ. «Για μένα, το [άστρο του Δαβίδ](#) δεν είναι μια μόδα που θα περάσει, αλλά
τη «περίθαψη» του πρώτου Ξηρού χωρίς [ένταλμα σύλληψης](#), ενώ λιγότερο δικαιολογημένα, δεδομένη
σε μηνιαία καταβολή μόνο τόκων.{S} Τα [έξοδα φακέλου](#) είναι μειωμένα κατά 50%.{S} Οι τράπεζες δ
ίνουν στο Λονδίνο μαζί με τα υπόλοιπα [έργα τέχνης](#)! {S}ΚΑΤΑΦΕΡΕ, έτσι, με έναν αστήρικτο και ψ
η...». {S}Ακούστηκε κι αυτό, καθώς τα [έργα υποδομής](#) συνεχίζονται στην πορεία των Ολυμπιακών Α
βεβαρημένο περιβάλλον οι εκσκαφές των [έργων υποδομής](#), αλλά και των ολυμπιακών κατασκευών δεν
</CATEGORY> <H2>«Είδαν» πλανήτη 5.000 [έτη φωτός](#) από τη Γη</H2> Με μια νέα τεχνική που θα χρησ

τη Νιγηρία. {S} Ο θεός πήγε να ψελίσει [ίχνη αλήθειας](#) και έφαγε τέτοια σφαλιάρα που του 'ρθε ο ιεργία πείνας</H3> Σταμάτησε χθες την [απεργία πείνας](#) που ξεκίνησε πριν από 23 μέρες η Αγγελικ του διαδραματίζει πρωτεύοντα ρόλο στην [απονομή δικαιοσύνης](#), ο ψυχικά άρρωστος κάτι κερδίζει κατάνου, ο οποίος θα ξεκινήσει για μια [αποστολή αυτοκτονίας](#). {S}Οι ζημιές Οι γερμανικές αρχές της Γεωργιάδης, για μία ληστεία, μία [απόπειρα ανθρωποκτονίας](#) και για κατοχή εκρηκτικών υλών φονία του αστυφύλακα Γιάννη Βαρή και [απόπειρα δολοφονίας](#) εις βάρος δώδεκα ακόμη αστυνομικών έθηκαν έπειτα από έρευνες του FBI στη [βάση δεδομένων](#) των επισκεπτών ενός site με παιδική πορν {S}Ο Ελαιώνας, που χαρακτηρίζεται... [γεφύρι της Άρτας](#), είναι περιοχή 9.400 στρεμμάτων. {S} Η να γίνονται τα πράγματα σύμφωνα με το [γράμμα του νόμου](#) και το πνεύμα του Συντάγματος. {S}Οι κ λευτούν έδιναν χρυσαφικά, τηλεκάρτες, [γυαλιά ηλίου](#) και φυσικά χρήματα, όπως αποδεικνύεται από για ζητήματα δικαιοσύνης στη βάση των [δικαιωμάτων του ανθρώπου](#). {S} Το αν επιθυμούμε, παντού σ θο, είναι μπροστάρης στη μάχη για την [ελευθερία της έκφρασης](#) στο ψηφιακό σύμπαν, το οποίο ο ί να σώσει αυτούς τους ανθρώπους από το [εμπόριο ναρκωτικών](#) και άλλες εγκληματικές δραστηριότητες, που «ξανασχεδίασαν» τον χάρτη της [εξέλιξης των ειδών](#). {S} Υποθέτουμε τις αιτίες τους, συνή τιο αξιολογούμε. {S} ότι η διάδοση του [επεξεργαστή κειμένου](#) κάνει ευκολότερη για τους συγγραφε γύρω από ένα άστρο σε απόσταση 5.000 [ετών φωτός](#), εντοπίζοντας μια ελαφρά αλλαγή στο φως που τίθεση. {S} Οι φιλοξενούμενοι με διπλή [ζώνη άμυνας](#) και με ασφυκτικό πρέσινγκ σε όλα τα μήκη κα να διατηρήσει τον μέσο πληθωρισμό στη [ζώνη του ευρώ](#) στο 2%, δηλαδή σε επίπεδο συνεπές με την στο παρά πέντε, στο σημείο μηδέν, στη [ζώνη του λυκόφωτος](#), τότε δεν τα μπουρδουκλώναμε όλα; {S} ή και απίθανα παραμύθια και γευθήκαμε [θεωρίες συνωμοσίας](#) για την «αληθινή ταυτότητα της 17N». ικαλούμενος κορυφαίος μορφές από την [ιστορία της τέχνης](#) αλλά και τον Καραγκιόζη, τον Θεόφιλο ιας και συγγραφικού ναρκισσισμού, μια [κίνηση εντυπωσιασμού](#) από συγγραφείς και εκδότες με βασι ζή για να παρακολουθήσει την εφαρμογή [κατάπαυσης του πυρός](#). {S} Ο ΟΗΕ έχει στείλει στην περιοχ ή πρωταγωνιστούν η αγάπη, ο φθόνος, η [κατάχρηση εξουσίας](#), η μαγεία, ο φόνος, η εκδίκηση, η λύ ; σε σχέση με αυτές που προσφέρουν οι [καταθέσεις ταμειευτηρίου](#). {S} Σύμφωνα με πληροφορίες, τα κατά τα άλλα οδηγεί τη χώρα μας στην "[κοινωνία της γνώσης](#)", είναι τεράστιες. {S} Η ανώτατη εκπ ριά τους, η κυρά-Κατερίνα, που 'ταν η [κολόνα του σπιτιού](#) αρρώστησε ακόμα πιο βαριά κι όλο το ως τα ευρήματα δεν αποτελούν παρά την [κορυφή του παγόβουνου](#). {S}Η ρικίνη. {S} Οι δηλητηριώδεις κι αγόραζε τέχνη χονδρικής, ότι έγινε [κυνηγός ταλέντων](#). {S} Κάποιοι απ' αυτούς χάθηκαν. {S} Έχε ; εξηγεί, πάντως, ο κ. Λαγουβάρδος το [κύμα κακοκαιρίας](#) που περίμεναν οι μετεωρολόγοι ήρθε πρά τρόσιτα για τη χώρα μας – υποτίθεται «[λίκνο του πολιτισμού](#)». {S}– Μην περιμένεις να έχει ακρο οσθέτει, «για να μειωθεί σημαντικά η [λίστα αναμονής](#) θα πρέπει κάθε τμήμα να έχει χειρουργείο ρηματοοικονομικά μέσα, όπως είναι οι [λογαριασμοί ταμειευτηρίου](#), οι οποίοι τους παρέχουν υπηρε η Πειραιώς, αλλά θα περιοριστούν οι [λωρίδες κυκλοφορίας](#) στους παρακηφίστους δρόμους. {S}Εντ > από το δικαστήριο ως κατηγοροι ή ως [μάρτυρες υπεράσπισης](#) κορυφαία στελέχη, κυρίως των δύο μ ακρίβεια, ακυρώνοντας κάθε συμβατικό [μέτρο ασφαλείας](#). {S} Μάλιστα καταλήγουν στο συμπέρασμα ό τιά της άλλους πέντε... {S}Βέβαια, τη [μερίδα του λέοντος](#) αυτό το πρώτο τρίμηνο – από Οκτώβριο αγόρι, προκειμένου να εξασφαλίσει το [μεροκάματο του τρόμου](#)! {S} Η συνέχεια είναι το «Κοπάδι» ρησιμοποιούν τον Έλληνα πλοίαρχο σαν [μοχλό πίεσης](#) προς την πλοιοκτήτρια εταιρεία, η οποία δε ις, αποκτήνωσης, βρώμικων συναλλαγών, [νόμου της ζούγκλας](#). {S}Οι φυλακισμένοι δεν είναι: 1. {S} εές εκλογές δεν μπορεί να βασιστεί σε [νόμους της σιωπής](#) και ζήτησε να γίνει βαθιά αυτοκριτική

is, που από το 1997 έχει αγοράσει τον [οίκο μόδας](#) του Τιερί Μιγκλέρ, τα μέλη του διοικητικού τ
 ύν, κυρίως γυναικών που εργάζονται σε [οίκους ανοχής](#) και προέρχονται από την Αλβανία.{S} Επίση
 εις-κόλαφους για τις συμπεριφορές των [οργάνων της τάξης](#) στα αστυνομικά τμήματα. <H3>ΝΙΚΟΣ ΧΡΙ
 ου Ν. 128/75).{S} Η δυνατότητα χρήσης [περιόδου χάριτος](#) κυμαίνεται από 6 έως 48 μήνες.{S} Σε μ
 Αθηνών. * Στη διαχείριση 3-4 μεγάλων [πνευμόνων πρασίνου](#) (Πεδίον του Άρεως, Αττικό Άλσος κ.λπ
 ορίς διαγωνισμό! {S}ΚΑΙ τα δύο, και η [ποιότητα ζωής](#) στις πόλεις αλλά και η εξυγίανση των δήμω
 αξίνωσης.{S} Και ετοιμάζεται να κάνει [πρόταση γάμου](#) στην νέα του αγαπημένη, ξεχνώντας, επίσης
 :ην αποκαλούσαν.{S} Ταυτιζόταν με μια [πόλη του φωτός](#), που η ανεξαρτησία της απειλείτο από την
 /, που επικοινωνήσε μαζί του μέσω του [πύργου ελέγχου](#).{S} Προσγειώθηκε αφού του επιτράπη να τη
 γτίο του πλοίου περιελάμβανε 437 κιλά [ράβδων χρυσού](#) σε 17 κασέλες, 15.399 χρυσά δουβλόνια, 15
 :μπόριο ή μια εξέγερση επηρεάζουν τον [ρου της Ιστορίας](#) και πως ο άνθρωπος πάντα μπορεί να προ
 εξακολουθεί να παραμένει πάνω από το [όριο ασφαλείας](#) και ήδη έχουν απομακρυνθεί από τις παρόχ

ii) Concordances du DELACF des *NI (E+DET:G) A:G N2:G* :

νθήματα εναντίον της Βορειοαμερικανικής [Ζώνης Ελευθέρων Συναλλαγών](#). «Κάντε έρωτα, όχι πόλεμο»,
 προς τον Σαντάμ Χουσεΐν. {S}Τα τουρκικά [Μέσα Μαζικής Ενημέρωσης](#) έδωσαν δημοσιότητα στην επίσκεψ
 με θέμα τα Ανθρώπινα Δικαιώματα και τα [Μέσα Μαζικής Ενημέρωσης](#). {S}Η έρευνα αφορούσε την παραβ
 αι δέχονται εύκολα σαν ελαφρυντικό τον «[βρασμό ψυχικής ορμής](#)» που σχεδόν πάντοτε επικαλείται ο
 μεγαλεμπόρων που στοκάρουν τρόφιμα και [είδη πρώτης ανάγκης](#) εν όψει της εισβολής στο Ιράκ...{S}
 ανική κυβέρνηση, συμβούλευε τους Ρώσους [ελεγκτές εναέριας κυκλοφορίας](#) που ετοιμάζονταν να κάνου
 κές ημέρες. {S}Επικαλέστηκε μάλιστα την [επιφοίτηση του Αγίου Πνεύματος](#) επί της πολιτικής ηγεσία
 και το σύστημα περιλαμβάνει εξέταση με [ερωτήσεις πολλαπλών επιλογών](#) από 5 βαθμολογικά ισότιμες
 τάσχουν σε «μια όσο το δυνατόν ευρύτερη [κυβέρνηση εθνικής ενότητας](#)», η οποία θα εξαντλήσει την
 άνετη πλειοψηφία.{S} Θα χαρακτηριζόταν [κυβέρνηση εθνικής ενότητας](#) και ο Σαρόν θα μπορούσε, θεω
 τρεμισμού» και ζητεί τη δημιουργία μιας [κυβέρνηση εθνικής ενότητας](#) με τη συμμετοχή των λαϊκών
 ο πλέον αξιόπιστο, ασφαλές και σύγχρονο [μέσο μαζικής μεταφοράς](#), του οποίου η επιβατική κίνηση α
 ευαισθητοποίηση των οικογενειών και των [μέσων μαζικής ενημέρωσης](#), που επίσης διαμορφώνουν τα πρ
 ίναι η μοναδική εταιρεία στον τομέα των [μέσων μαζικής μεταφοράς](#), που όχι μόνο δεν επιδοτείται α
 υ διατηρεί ο ίδιος στην πόλη. {S}Από το [πρατήριο υγρών καυσίμων](#) ιδιοκτησίας του Παπαδόπουλου πρ
 , πίσω από τις κάμερες διαδραματίζονταν [σκηνές απείρου κάλλους](#) με τις... συμμορίες του Λος Άντζ
 έλλας Θεοδωράκη (τέλη Μαρτίου), η πρώτη [ταινία μεγάλου μήκους](#) της σκηνοθέτιδος «με τον ανέφικτο
 ίων.{S} Έτσι, στα περισσότερα από 5.000 [φροντιστήρια ξένων γλωσσών](#), που λειτουργούν υπό την αιγ
 λευταίων Ημερών». {S} Παρότι Εκκλησία το [φροντιστήριο ξένων γλωσσών](#), βρίσκω σατανικό κυριολεκτικ
 γέτη, ούτε καν για τον εγκλεισμό του σε [φυλακή υψίστης ασφαλείας](#) της Τουρκίας.{S} Το πέρασμά το
 S} Ήταν κατάδικος στην πιο εκλεπτυσμένη [φυλακή υψίστης ασφαλείας](#) του κόσμου», λέει ο Νόρμαν Μέι
 που περιείχε το αγγείο. <H3>Φιλέτα από [ψάρια του γλυκού νερού](#)!</H3> «Το ερώτημα που προέκυψε ή

11.3.2 Les règles de grammaire

Une autre méthode pour la reconnaissance lexicale automatique consiste à utiliser des règles
 de grammaire exprimées sous forme de graphes d'automates à états finis, d'expressions

régulières ou bien de règles construites *ad hoc*. Dans le cadre de la présente étude, nous avons opté pour la construction de graphes d'automates à états finis à l'aide d'Unitex (S. Paumier 2006). Cet outil informatique offre plusieurs possibilités de manipulation de données à l'aide de ce formalisme, à savoir : construction manuelle de grammaires locales composées de graphes d'automates finis, conversion semi-automatique des tables du lexique-grammaire en graphes d'automates finis par le biais des graphes paramétrés, étiquetage des unités lexicales reconnues grâce aux transducteurs).

« Les grammaires locales sont un moyen puissant de représenter la plupart des phénomènes linguistiques » (S. Paumier 2006 : 65). Les grammaires que nous utilisons ont donc la forme de réseaux récurrents de transition (RTN). Les graphes se lisent de gauche à droite. Ils comportent des chemins qui partent d'un état initial et aboutissent à un état final. Chaque chemin est étiqueté par des mots, des chiffres ou des symboles. Un chemin représente une suite de ces derniers qui est une unité du lexique ou/et de la grammaire. La notation $\langle E \rangle$ est utilisée pour la séquence vide. Dans ces graphes, on se donne la possibilité d'insérer des nœuds auxiliaires (appelés sous-graphes), c'est-à-dire des nœuds qui sont eux-mêmes des RTN (M. Silberstein 1993 : 21-30). De cette manière, nous évitons des descriptions redondantes. Il serait, d'ailleurs, impossible de les décrire « à la main » avec les seuls graphes lexicalisés. Enfin, les graphes d'automates permettent l'introduction de variables quelconques.

« Les graphes peuvent être assimilés à des extensions des dictionnaires des mots composés » (M. Constant 2003 : 36). Nous utilisons cette méthode pour la reconnaissance automatique des syntagmes nominaux semi-figés (ou productifs) de structure $N(E+DET:G)N:G$ qui ont un paradigme ouvert.

Rappelons que les entités nommées ne sont pas traitées dans le cadre de la présente étude. Les graphes qui servent à reconnaître les entités nommées sont différents des graphes que nous utilisons pour le traitement des syntagmes nominaux productifs (ou semi-figés), en raison de leurs fonctionnalités spécifiques (prise en compte du contexte immédiat droit ou gauche, exploitation du contexte éloigné, utilisation de variables etc.) (cf. C. Martineau et al. 2007).

A titre indicatif, nous présentons ci-après un exemple d'un extrait de la grammaire locale que nous avons construite pour les $N(E+DET:G)N:G$ productifs (ou semi-figés) dont le nom-tête est *κέντρο/centre*, ainsi qu'un extrait des résultats obtenus après son application :

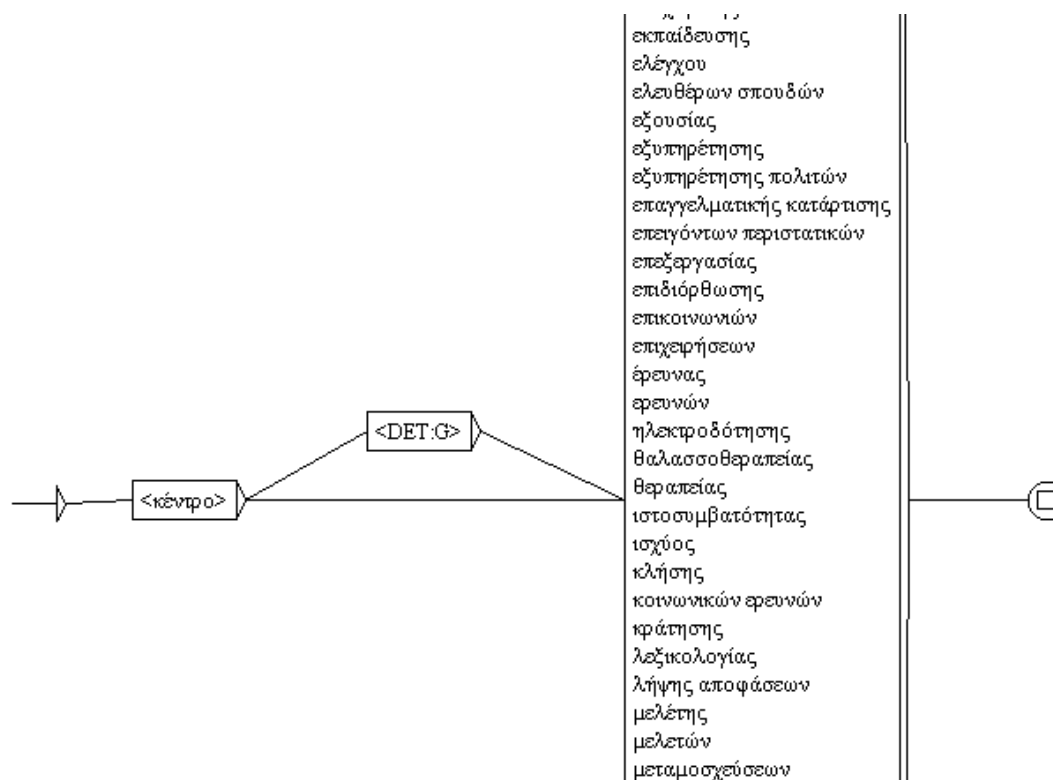


Figure 5 : Exemple de graphe lexicalisé

τά της γραφειοκρατίας θα αποτελέσουν τα [Κέντρα Εξυπηρέτησης Πολιτών](#).{S} Μετά τη δυνατότητα υποβ
α τον σκοπό αυτό: * Ιδρύονται άμεσα δύο [Κέντρα Εξυπηρέτησης Πολιτών](#) (Αθήνα – Πειραιά), ενώ από
ς μετανάστριας, όπως καταγράφηκε από το [Κέντρο Ερευνών](#) για Θέματα Ισότητας (ΚΕΘΙ), σύμφωνα με τ
ενάμιση χρόνο η Δήμητρα απευθύνθηκε στο [Κέντρο Ερευνών](#) «Ρίζες».{S} Την υπόθεσή της ανέλαβε η Μα
πολιτιστικής παράδοσης», Εκδ.{S} Εθνικό [Κέντρο Κοινωνικών Ερευνών](#), τιμή: {S}17 ευρώ. <REF> <SOU
γίας και διευθύντρια ερευνών στο Εθνικό [Κέντρο Κοινωνικών Ερευνών](#), δηλώνει «ταγμένη ερευνήτρια»
εριμαζέψει και να τα οδηγήσει σε κάποιο [Κέντρο Περίθαλψης](#), όπως όφειλε σύμφωνα με τον νόμο. {S}
Σακελλάριο. <H3>INFO</H3> Η έκθεση στο [Κέντρο Σύγχρονης Τέχνης](#) ΟΜΜΑ των Χανίων (Ελ.{S} Βενιζέλ
μη, ενώ οι άλλοι τρεις νοσηλεύονται στο [Κέντρο Υγείας](#) Καστελλίου. {S}Κατολισθήσεις στο Πήλιο Απ
ωτερικό.{S} Η λειτουργία ενός σύγχρονου [Κέντρου Διαχείρισης](#) Κυκλοφορίας, η δημιουργία σταθμών μ
ς Οικονομίας και γενικός διευθυντής του [Κέντρου Ερευνών](#) και Ανάπτυξης – Intercollege. <REF> <SO
το Πολιτικής Κοινωνιολογίας του Εθνικού [Κέντρου Κοινωνικών Ερευνών](#) <REF> <SOURCE>TA NEA</SOURCE>
το Πολιτικής Κοινωνιολογίας του Εθνικού [Κέντρου Κοινωνικών Ερευνών](#) <REF> <SOURCE>TA NEA</SOURCE>
ώτη (της Νέας Ελληνικής Γλώσσας, έκδοση [Κέντρου Λεξικολογίας](#)) αναφέρει στο σχετικό λήμμα τρεις
ρινούς μήνες, είχαν και οι εκθέσεις του [Κέντρου Σύγχρονης Τέχνης](#) Ρεθύμνης, στη «Φορτέτσα» (χώρο
της Πανελληνίας Ομοσπονδίας Ιδιοκτητών [Κέντρων Ξένων Γλωσσών](#), παρακολουθούν μαθήματα περίπου έ

ρεπε να δημιουργηθούν δύο ανταγωνιστικά κέντρα εξουσίας και μάλιστα στην ίδια πόλη, γιατί αυτό τηρείται και το μεγαλύτερο πρόβλημα, τα κέντρα εξυπηρέτησης θα αυξηθούν από τρία σε οκτώ Κάτοχο ρχιών. {S}Ειδικά στη Νομαρχία Αθηνών τα κέντρα εξυπηρέτησης μεταναστών θα αυξηθούν από 3 σε 8, ς καλά κρατεί...</H2> <H4>Οι άνδρες στα κέντρα λήψης αποφάσεων, οι γυναίκες στην οικογένεια...{ ορήθηκαν ότι μετέτρεψε το σπίτι της σε «κέντρο διανομής» και ακολούθησαν δύο τοξικομανείς, το β γκαράζ, δεξαμενές νερού, καθώς και ένα κέντρο διαχείρισης αποβλήτων. {S}Σύμφωνα με το «Wired O υθρες ακτίνες, που θα συνδέονται με ένα κέντρο επιχειρήσεων, το οποίο θα ελέγχεται από κοινού, ελική του μορφή το έργο θα περιλαμβάνει κέντρο πληροφόρησης, μουσείο φυσικής ιστορίας, προβλήτα ύ, περιμένουν στην ουρά για βοήθεια στο κέντρο προσφύγων της Μανγκίνα «Τους είδα να κόβουν τον ίες αποδεικνύουν πως ήταν και προηγμένο κέντρο υφαντουργίας. <H3>Τόπος τελειών</H3> Ενδιαφέρον οίκο ανοχής και πλούτισε ως ιδιοκτήτης κέντρων διασκέδασης, η Γκάλους Μαγκ (Μαγκ η Τιράντω), δ ιβασμούς στο εσωτερικό της διαφορετικών κέντρων ισχύος, διαφορετικών αντιλήψεων και ιεραρχήσεων ερα – θα επιβεβαιώσουν οι υπεύθυνες των κέντρων ομορφιάς και SPA στα προάστια της Αθήνας – κλεί

Les résultats ne sont pas très satisfaisants. Dans la plupart des cas, la forme obtenue fait partie d'une forme plus longue qui est souvent une entité nommée du type « organisation ». Par exemple, nous obtenons *Κέντρο Ερευνών*/Centre de recherches, alors que la forme complète est *Κέντρο Ερευνών για Θέματα Ισότητας*/Centre de recherches **pour des questions d'égalité**. Nous observons que souvent la forme identifiée est suivie d'un toponyme, e.g. *Κέντρο Υγείας Καστελλίου*/Centre de santé **de Kastelli**, *Κέντρο Σύγχρονης Τέχνης Ρεθύμνης*/Centre d'art moderne **de Rethymna**, *κέντρο προσφύγων της Μανγκίνα*/centre de réfugiés **de Manguina**. Enfin, dans certains cas, la forme reconnue est erronée car incomplète, e.g. *κέντρο διαχείρισης*/centre de gestion au lieu de *κέντρο διαχείρισης αποβλήτων*/centre de gestion **des déchets**.

Notons aussi qu'un autre problème des graphes est leur nombre trop important et l'impossibilité de compactage des données décrites dans ceux-ci. L'impossibilité de compactage complique considérablement leur utilisation et leur application dans les textes.

En l'absence de solution plus satisfaisante, nous avons opté pour la solution provisoire de la construction de graphes lexicalisés pour les $N (E+DET:G) N:G$ productifs (ou semi-figés). Un extrait représentatif de ces graphes est présenté en Annexe 5.

11.3.3 Les tables du lexique-grammaire

Les tables du lexique-grammaire sont un moyen compact de décrire systématiquement et représenter formellement les propriétés syntaxico-sémantiques des éléments lexicaux d'une langue. Ces descriptions systématiques sont représentées au moyen de matrices binaires où les

lignes correspondent à des entrées lexicales (dans notre cas, à des $N (E+DET:G) N:G$ binaires ou surcomposés) et les colonnes aux propriétés de ces entrées. Chaque table est saisie à l'aide d'un tableur, comme celui de Microsoft Excel. Une cellule de la table (à l'intersection d'une ligne et d'une colonne) contient soit du texte, soit un signe « + » ou « - ». Ce codage apparaît pour préciser si l'élément lexical de cette ligne vérifie ou non la propriété de cette colonne.

Le format de la table doit respecter les contraintes que nous résumons ci-après pour qu'elle puisse être utilisée à des fins de traitement automatique des textes :

- la première ligne contient le nom de chaque zone de texte ou de propriété (voire, les en-têtes de chaque colonne). Il doit y avoir autant d'en-têtes que de colonnes ;
- une colonne de texte contient une séquence qui peut comporter soit du texte (*i.e.* $\nu\delta\rho\omicron\gamma\omicron\nu\omicron\beta\omicron\mu\beta\alpha$ /bombe à hydrogène) soit le symbole <E> pour représenter la séquence vide ;
- une colonne de propriété contient exclusivement un signe « + » ou « - » ;
- les colonnes doivent être homogènes ; par exemple, une colonne de propriété ne peut contenir aucun texte.

Dans le cadre de cette étude, nous avons construit deux tables du lexique-grammaire, notamment dans le but d'effectuer la mise en relation des variantes syntaxiques des $N (E+DET:G) N:G$ et des $NI (E+DET:G) A:G N2:G$. Précisons que ces tables ne correspondent pas à des phrases simples.

A titre d'illustration, nous présentons ci-après des extraits de ces tables du lexique-grammaire accompagnés chacun d'un tableau qui fournit des explications sur les propriétés qui y figurent. Notons que des extraits représentatifs de nos tables du lexique-grammaire figurent en Annexes 3 et 4.

Soulignons que, à l'heure actuelle, nous nous sommes servie des tables du lexique-grammaire pour la représentation formalisée des variantes syntaxiques ainsi que pour l'ajout des traits sémantiques et du trait syntaxico-sémantique *Npréd*, car il n'était pas possible de fournir ces informations dans le dictionnaire électronique morphologique. Nous envisageons de compléter également l'élaboration de notre dictionnaire électronique morphologique en y ajoutant toutes les informations précitées.

i) La table du lexique-grammaire pour les $N (E+DET:G) N:G$ du grec moderne :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
N	DET:G	N:G	N DET:G N:G	N N:G	DET:G=E	DET:G N:G	Namb	N DET:G N:G=N:G-a N		N:G-a	N DET:G N:G=Ns	Ns	N DET:G N:G=Npréd	N DET:G N:G=Nhum	N DET:G N:G=Nhumcoll	N DET:G N:G=Nanim	N DET:G N:G=Nconc	N DET:G N:G=Nabs	N DET:G N:G=TS
ανακοπή	της	καρδιάς	+	-	+	-	+	-	<E>	-	<E>	+	-	-	-	-	+	+	
ανίχνευση	του	πυρός	+	-	-	-	-	-	<E>	+	πυρανίχνευση	+	-	-	-	-	+	-	
ανιχνευτής	-	ψεύδους	-	+	-	-	-	-	<E>	-	<E>	-	-	-	-	-	+	-	
βαλβίδα	-	ασφαλείας	-	+	-	-	-	+	ασφαλιστικός.A:f	-	<E>	-	-	-	-	+	-	-	
βάπτισμα	του	πυρός	+	-	-	-	-	-	<E>	-	<E>	-	-	-	-	-	+	-	
βόμβα	-	υδρογόνου	-	+	-	-	-	-	<E>	+	υδρογονοβόμβα	-	-	-	-	+	-	-	
βούτυρο	του	κακάο	+	-	+	-	-	-	<E>	-	<E>	-	-	-	-	+	-	-	
διάσειση	του	εγκεφάλου	+	-	+	-	-	+	εγκεφαλικός.A:f	-	<E>	-	-	-	-	+	-	+	
δίνη	-	ανέμου	-	+	-	-	-	-	<E>	+	ανεμοστρόβιλος	-	-	-	-	+	-	-	
επιστροφή	του	φόρου	+	-	+	-	-	-	<E>	-	<E>	+	-	-	-	-	+	-	
ίνα	-	γυαλιού	-	+	-	-	-	+	γυάλινος.A:f	-	<E>	-	-	-	-	+	-	-	
ίππος	-	άλματος	-	+	-	-	-	-	<E>	-	<E>	-	-	-	-	+	-	-	
καφές	-	στιγμής	-	+	-	-	-	+	στιγμιαίος.A:f	-	<E>	-	-	-	-	+	-	-	
κενό	-	αέρος	-	+	-	-	+	-	<E>	-	<E>	-	-	-	-	+	-	-	
λάδι	της	ελιάς	+	-	+	-	-	-	<E>	+	ελαιόλαδο	-	-	-	-	+	-	-	

Figure 7 : Extrait de la table du lexique-grammaire pour les $N (E+DET:G) N:G$

Pour faciliter la lecture de cette table, nous présentons ci-après un tableau, qui récapitule tous les intitulés des colonnes, leur explication et un exemple :

Intitulé de la colonne	Explication	Exemple
N	Propriété à valeur lexicale. Le nom-tête du $N (E+DET:G) N:G$.	φακός επαφής/ lentille de contact
$DET:G$	Propriété à valeur lexicale. Le déterminant au génitif (optionnel) du $N (E+DET:G) N:G$.	πίατο (E + της) ημέρας/ plat du jour
$N:G$	Propriété à valeur lexicale. Le deuxième constituant du $N (E+DET:G) N:G$, toujours au génitif.	φακός επαφής / lentille de contact
$N DET:G N:G$	Propriété à valeur binaire. Correspond à la construction de base avec déterminant.	πίατο της ημέρας / plat du jour
$N N:G$	Propriété à valeur binaire.	πίατο ημέρας / plat du jour

	Correspond à la construction de base sans déterminant.	
DET:G=E	Propriété syntaxique à valeur binaire. Réduction du déterminant.	<i>πίατο (E + της) ημέρας</i> /plat du jour
DET:G N:G	Propriété syntaxique à valeur binaire. Réduction du nom-tête.	<i>(E+εισιτήριο)</i> <i>διαρκείας</i> /carte d'abonnement
Namb	Propriété syntaxique à valeur binaire. Réduction du déterminant et du deuxième constituant. La variante réduite ainsi obtenue est un nom ambigu.	<i>δίπλωμα</i> <i>(E+οδήγησης)</i> /permis de conduire
N DET:G N:G=N:G-a N	Propriété syntaxique à valeur binaire. Mise en relation du <i>N (E+DET:G) N:G</i> avec la variante de type AN.	<i>σύνδρομο</i> <i>στέρησης=στερητικό</i> <i>σύνδρομο</i> /syndrome de sevrage
N:G-a N	Propriété à valeur lexicale. L'adjectif morphologiquement associé au deuxième constituant du <i>N (E+DET:G) N:G</i> .	<i>στερητικός</i> /de sevrage
N DET:G N:G=Ns	Propriété syntaxique à valeur binaire. Mise en relation du <i>N (E+DET:G) N:G</i> avec la variante soudée.	<i>σίδερο</i> <i>ατμού=ατμοσίδερο</i> /fer à repasser vapeur
Ns	Propriété à valeur lexicale. Le nom soudé, qui constitue la variante du <i>N (E+DET:G) N:G</i> .	<i>ατμοσίδερο</i> /fer à repasser vapeur
N DET:G N:G=Npred	Propriété syntaxique à valeur binaire. Le <i>N (E+DET:G) N:G</i> est nom prédicatif.	<i>απεργία πείνας</i> /grève de faim
N DET:G N:G=:Nhum	Propriété sémantique à valeur binaire. Le <i>N (E+DET:G) N:G</i> est un nom humain.	<i>εγκληματίας πολέμου</i> /criminel de guerre

<i>N DET:G N:G=:Nhumcoll</i>	Propriété sémantique à valeur binaire. Le $N (E+DET:G) N:G$ est un nom humain collectif.	άσυλο απόρων/maison de retraite
<i>N DET:G N:G=:Nanim</i>	Propriété sémantique à valeur binaire. Le $N (E+DET:G) N:G$ est un nom animé	γάτα Αγκύρας/chat angora
<i>N DET:G N:G=:Nconc</i>	Propriété sémantique à valeur binaire. Le $N (E+DET:G) N:G$ est un nom concret.	ανιχνευτής ψεύδους/détecteur de mensonges
<i>N DET:G N:G=:Nabs</i>	Propriété sémantique à valeur binaire. Le $N (E+DET:G) N:G$ est un nom abstrait.	απεργία πείνας/grève de faim
<i>N DET:G N:G=:TS</i>	Propriété sémantique à valeur binaire. Le $N (E+DET:G) N:G$ est un terme de spécialité.	νόσφιση εξουσίας/usurpation de pouvoir

ii) La table du lexique-grammaire pour les *N1 (E+DET:G) A:G N2:G* du grec moderne :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
N1	DET:G	A:G	N2:G	N1 A:G N2:G	N1 DET:G N2:G	A:G N2:G	N1 DET:G A:G	N1	N1 DET:G A:G N2:G=<A:G N2:G>=a N1	<A:G N2:G>=a	N1 DET:G A:G N2:G=A N1 DET:G N2:G	N1 DET:G A:G N2:G=A N1	N1 DET:G A:G N2:G=A N2	N2	A	N1 DET:G A:G N2:G = A:G N2:G N1	N1 DET:G A:G N2:G=Npréd	N1 DET:G A:G N2:G=Nhum	N1 DET:G A:G N2:G=Nhumcoll	N1 DET:G A:G N2:G=Nanimé	N1 DET:G A:G N2:G=Nconc	N1 DET:G A:G N2:G=Nabs	N1 DET:G A:G N2:G=TS
αγγελία	-	μέλλοντος	γάμου	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	+	-	-	-	-	+	-
αγγελία	-	μέλλοντος	γάμου	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	-	+	-	-
άγγελος	-	κακών	ειδήσεων	-	-	-	+	-	-	<E>	-	-	-	<E>	<E>	-	-	+	-	-	-	-	-
άγγελος	-	κακών	επών	-	-	-	+	-	-	<E>	-	-	-	<E>	<E>	-	-	+	-	-	-	-	-
άγγελος	-	καλών	ειδήσεων	-	-	-	+	-	-	<E>	-	-	-	<E>	<E>	-	-	+	-	-	-	-	-
αγαγός	-	όμβριων	υδάτων	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	+	-	-	-
αγαγός	-	ψυσικού	αερίου	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	+	-	-	-
αγώνας	-	ανώμαλου	δρόμου	-	-	-	-	-	-	<E>	-	-	+	<δρόμος.N>	<ανώμαλος>	-	+	-	-	-	-	+	+
αίθουσα	-	πολλαπλών	χρήσεων	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	+	-	-	-
αιχμή	-	ηλεκτρικού	ρεύματος	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	-	+	+	-
αιχμή	-	ηλεκτρικού	φορτίου	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	-	+	+	-
αλκυόνι	-	αλκυόνος	άλκυονος	-	-	-	-	-	-	<E>	-	-	-	<E>	<E>	-	-	-	-	+	-	-	-

Figure 8 : Extrait de la table du lexique-grammaire pour les *N1 (E+DET:G) A:G N2:G*

Ci-après un tableau, qui récapitule tous les intitulés des colonnes, leur explication et un exemple :

Intitulé de la colonne	Explication	Exemple
<i>N1</i>	Propriété à valeur lexicale. Le nom-tête du <i>N1 (E+DET:G) A:G N2:G</i> .	<i>οικονομία</i> (<i>E+της</i>) (<i>E+ελεύθερης</i>) <i>αγοράς/économie</i> du marché
<i>DET:G</i>	Propriété à valeur lexicale. Le déterminant au génitif (optionnel) du <i>N1 (E+DET:G) A:G N2:G</i> .	<i>οικονομία της</i> (<i>E+ελεύθερης</i>) <i>αγοράς/économie</i> du marché
<i>A:G</i>	Propriété à valeur lexicale. L'adjectif du <i>N1 (E+DET:G) A:G N2:G</i> au génitif.	<i>οικονομία</i> (<i>E+της</i>) <i>ελεύθερης</i> <i>αγοράς/économie</i> du marché
<i>N2:G</i>	Propriété à valeur lexicale. Le deuxième complément nominal du	<i>οικονομία</i> (<i>E+της</i>) <i>ελεύθερης</i> <i>αγοράς/économie</i> du marché

	<i>NI (E+DET:G) A:G N2:G</i> , toujours au génitif.	
<i>NI A:G N2:G</i>	Propriété syntaxique à valeur binaire. Réduction du déterminant..	<i>οικονομία ελεύθερης αγοράς</i> /économie du marché
<i>NI DET:G N2:G</i>	Propriété syntaxique à valeur binaire. Réduction du modifieur adjectival.	<i>οικονομία αγοράς</i> /économie du marché
<i>A:G N2:G</i>	Propriété syntaxique à valeur binaire. Réduction du nom-tête.	<i>ελευθέρως εισόδου</i> /bulletin d'accès libre
<i>NI DET:G A:G</i>	Propriété syntaxique à valeur binaire. Réduction du deuxième complément nominal et substantivation de l'adjectif.	<i>άγγελος κακών (Ε + επών)</i> /porteur de mauvaises nouvelles
<i>NI</i>	Propriété syntaxique à valeur binaire. Réduction de l'adjectif et du deuxième complément nominal.	<i>δίκτυο (Ε+ηλεκτρονικών υπολογιστών)</i> /réseau (E+d'ordinateurs)
<i>NI DET:G A:G N2:G= <A:G N2:G>-a NI</i>	Propriété syntaxique à valeur lexicale. Adjectivation du <i>A:G N2:G</i> par la formation d'un adjectif composé soudé.	<i>έργο κοινής ωφελείας=κοινοφελές έργο</i> /travail d'utilité publique
<i><A:G N2:G>-a</i>	Propriété à valeur lexicale. L'adjectif composé soudé.	<i>κοινοφελές</i> /d'utilité publique
<i>NI DET:G A:G N2:G= A NI DET:G N2:G</i>	Propriété syntaxique à valeur binaire. L'adjectif du deuxième complément nominal est antéposé.	<i>μέσο μαζικής μεταφοράς=μαζικό μέσο μεταφοράς</i> /moyen de transport
<i>NI DET:G A:G N2:G= A NI</i>	Propriété syntaxique à valeur binaire. Antéposition de l'adjectif et effacement du deuxième complément nominal.	<i>επιχείρηση μικτής οικονομίας=μικτή επιχείρηση</i> /entreprise d'économie mixte
<i>NI DET:G A:G N2:G= A N2</i>	Propriété syntaxique à valeur binaire. Réduction du nom-tête et formation d'un nouveau AN.	<i>δελτίο αστυνομικής ταυτότητας=αστυνομική ταυτότητα</i> /carte d'identité
<i>N2</i>	Propriété syntaxique à valeur binaire.	<i>ταυτότητα</i> /carte d'identité

	Réduction du nom-tête et du modifieur adjectival.	
<i>A</i>	Propriété syntaxique à valeur binaire. Réduction du nom-tête et du deuxième complément nominal.	<i>δελτίο</i> <i>ελευθέρας</i> <i>εισόδου</i> /bulletin d'accès libre
<i>N1 DET:G A:G N2:G = A:G N2:G N1</i>	Propriété syntaxique à valeur binaire. Permutation.	<i>κάθειρξη</i> <i>αόριστης</i> <i>διάρκειας</i> = <i>αόριστης</i> <i>διάρκειας</i> <i>κάθειρξη</i> / emprisonnement à durée indéterminée
<i>N1 DET:G A:G N2:G=Npred</i>	Propriété syntaxique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est nom prédicatif.	<i>αγγελία</i> <i>μέλλοντος</i> <i>γάμου</i> /annonce de mariage
<i>N1 DET:G A:G N2:G=:Nhum</i>	Propriété sémantique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est un nom humain.	<i>άγγελος κακών (E + επών)</i> /porteur de mauvaises nouvelles
<i>N1 DET:G A:G N2:G=:Nhumcoll</i>	Propriété sémantique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est un nom humain collectif.	<i>επιχείρηση μικτής οικονομίας</i> /entreprise d'économie mixte
<i>N1 DET:G A:G N2:G=:Nanim</i>	Propriété sémantique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est un nom animé	
<i>N1 DET:G A:G N2:G=:Nconc</i>	Propriété sémantique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est un nom concret.	<i>γάλα μακράς διαρκείας</i> /lait à longue durée de conservation
<i>N1 DET:G A:G N2:G=:Nabs</i>	Propriété sémantique à valeur binaire. Le <i>N1 (E+DET:G) A:G N2:G</i> est un nom abstrait.	<i>αγγελία</i> <i>μέλλοντος</i> <i>γάμου</i> /annonce de mariage

<i>NI</i> <i>DET:G</i> <i>A:G</i> <i>N2:G=TS</i>	Propriété sémantique à valeur binaire. Le <i>NI (E+DET:G) A:G N2:G</i> est un terme de spécialité.	<i>δέσμη</i> <i>φωτεινών</i> <i>ακτίνων</i> /faisceau lumineux
---	---	---

11.3.3.1 Conversion automatique des tables en grammaires

Une fois terminée la construction des tables du lexique-grammaire pour les *N (E+DET:G) N:G* et les *NI (E+DET:G) A:G N2:G*, on doit transformer automatiquement ces tables en graphes d'automates à états finis à l'aide d'un graphe paramétré (cf. V, 11.3.3.2) afin qu'elles puissent être appliquées aux textes grecs. Le graphe paramétré génère automatiquement, à partir d'une table, le graphe spécialisé pour chacune de ses lignes (ou entrées lexicales). L'ensemble des graphes générés constitue une grammaire locale qui sert à identifier les constructions décrites dans les tables.

Pour effectuer ce travail, il nous faut, pour chacune des tables du lexique-grammaire :

- le graphe paramétré qui formalise les propriétés décrites dans cette table ;
- un outil informatique (en l'occurrence Unitex) qui met en relation les propriétés dans les tables et les chemins des graphes correspondants, et qui produit des transducteurs. Les transducteurs sont des graphes qui associent à des séquences reconnues des informations produites (cf. M. Silberstein 1993).

11.3.3.2 Les graphes paramétrés

La conversion d'une table en grammaire s'effectue au moyen du mécanisme des graphes paramétrés. Le graphe paramétré d'une table du lexique-grammaire est utilisé pour interpréter les propriétés décrites dans cette table, puisque chacune de ses entrées est confrontée automatiquement au graphe paramétré. Un graphe paramétré se différencie d'un simple graphe lexical par les références à des cellules.

Plus explicitement, chaque construction, qui figure en colonne, est représentée par un chemin au sein du graphe. Le graphe fait référence aux colonnes grâce à des variables (introduites par le symbole '@'). Par exemple, @A renvoie à la première colonne de la table, @B renvoie à la deuxième colonne, et ainsi de suite. La valeur de cette variable dépend directement du contenu de l'entrée lexicale concernée. Ainsi, si la cellule contient le signe « + », le chemin

est retenu. Si elle contient le signe « - », le chemin est supprimé du graphe. Enfin, lorsque la cellule contient du texte, la variable est remplacée par ce contenu.

11.3.3.3 Application aux $N (E+DET:G) N:G$

Dans le cadre de cette étude, nous avons appliqué la méthode décrite ci-dessus sur les données lexicales figurant dans la table du lexique-grammaire des $N (E+DET:G) N:G$.

Ainsi, à partir de cette table (cf. Figure 7), nous avons construit le graphe paramétré qui formalise les propriétés décrites dans celle-ci. Nous présentons ci-après un extrait du graphe paramétré :

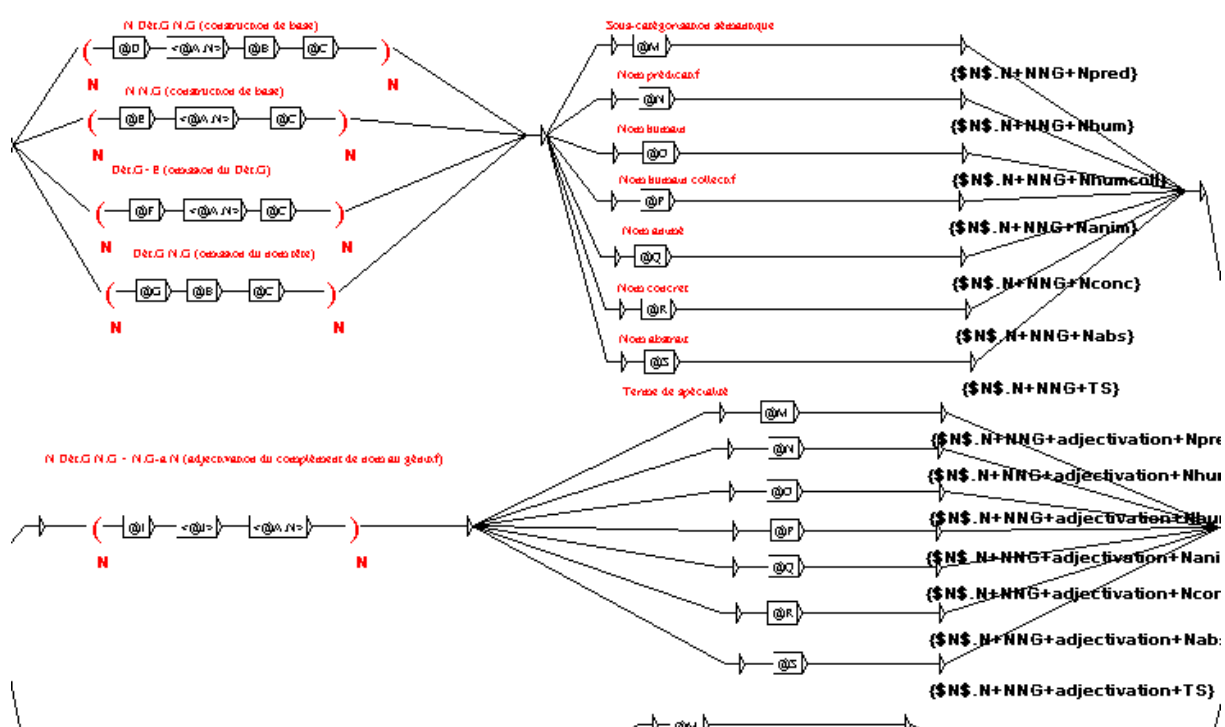


Figure 9 : extrait du graphe paramétré

Dans cet extrait du graphe :

- les variables '@E', '@F', etc. renvoient aux identifiants des colonnes de la table correspondante (cf. Figure 7) ;
- est représentée la possibilité de réduction du $DET:G$ (troisième chemin, colonne F) ;

- est représentée la possibilité de réduction du nom-tête (quatrième chemin, colonne G) ;
- est représentée la possibilité d'adjectivation du complément de nom au génitif (cinquième chemin, colonnes I, J) ;
- les variables '@M'-'@S' renvoient aux identifiants des colonnes de la table correspondant à la sous-catégorisation sémantique que nous avons établie (nom humain, nom abstrait etc.) (cf. IV, 10).

A partir du graphe paramétré, un graphe principal est généré automatiquement qui fait appel à des sous-graphes. 2 547 sous-graphes sont générés et correspondent aux 2 547 entrées figurant dans la table du lexique-grammaire. Nous présentons ci-après un extrait de notre graphe principal appelant tous les graphes générés :

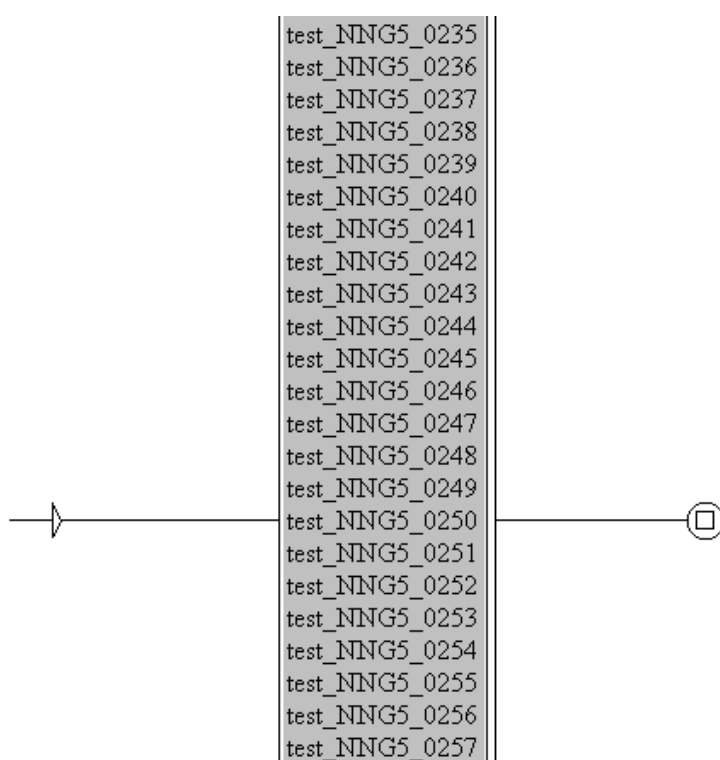


Figure 10 : Extrait du graphe principal appelant tous les graphes générés

Nous présentons ci-après le sous-graphe généré pour l'entrée *σύνδρομο (Ε+της) στέρησης/syndrome de sevrage*.

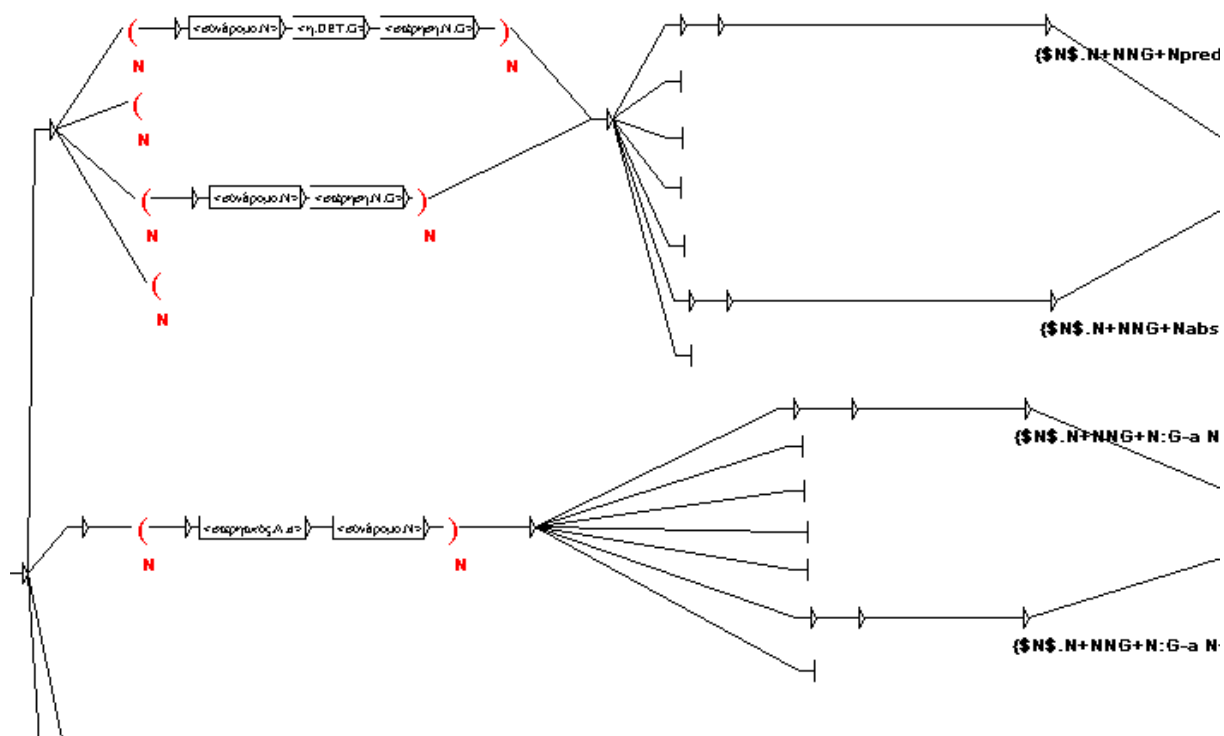


Figure 11 : Extrait de sous-graphe

Ce graphe permet de reconnaître :

- la forme canonique, dite structure de base (premier chemin) ;
- la variante $N N:G$ obtenue après réduction du $DET:G$ (deuxième chemin) ;
- la variante AN issue de l'adjectivation du complément de nom au génitif (troisième chemin).

L'application du graphe en mode « merge » permet d'attribuer à l'ensemble des formes reconnues leur catégorie grammaticale, leur sous-catégorie morphosyntaxique ainsi que les traits sémantiques appropriés.

Nous avons appliqué l'ensemble des graphes générés au même corpus que le dictionnaire électronique morphologique (journal TA NEA, janvier 2003). A titre indicatif, nous présentons un extrait des résultats obtenus après leur application :

νισμένο * 250 γρ. κεφαλοτύρι τριμμένο * [Αλάτι {Αλάτι.N+NNG+ambigu+Nconc}](#) και πιπέρι * 1 κούπα ελ
του Στίβεν Σόντερς, στις 8* 6* 2000. * [Απόπειρα ανθρωποκτονίας {Απόπειρα ανθρωποκτονίας.N+NNG+Nabs}](#)
του Στίβεν Σόντερς, στις 8* 6* 2000. * [Απόπειρα ανθρωποκτονίας {Απόπειρα ανθρωποκτονίας.N+NNG+Npred}](#)

ών εις βάρος πέντε διερχομένων πεζών. * [Απόπειρα δολοφονίας{Απόπειρα δολοφονίας.N+NNG+Nabs}](#) από
 ών εις βάρος πέντε διερχομένων πεζών. * [Απόπειρα δολοφονίας{Απόπειρα δολοφονίας.N+NNG+Npred}](#) απ
 λακτικά, σαμπρέλες. {S}Τι ήθελε αυτός ο [άγγελος{άγγελος.N+NNG+ambigu+Nhum}](#) ανάμεσά τους;{S} Προ
 νεται στον Δαρείο που στέκεται πάνω στο [άρμα{άρμα.N+NNG+ambigu+Nconc}](#) του, στη μάχη της Ισοού.{
 νάμεων με περίπου 4.000 άνδρες, και 120 [άρματα μάχης{άρματα μάχης.N+NNG+Nconc}](#) λόγω της πιθανής
 των Ηνωμένων Πολιτειών κι από πίσω του [άρματα μάχης{άρματα μάχης.N+NNG+Nconc}](#) που κάνουν ασκήσ
 δρόμους από όπου περνούσαν τα ισραηλινά [άρματα{άρματα.N+NNG+ambigu+Nconc}](#) εκρηκτικούς μηχανισμό
 φορά απόρριψη με μυστική ψηφοφορία της [άρσης ασυλίας{άρσης ασυλίας.N+NNG+Nabs}](#) επτά βουλευτών,
 φορά απόρριψη με μυστική ψηφοφορία της [άρσης ασυλίας{άρσης ασυλίας.N+NNG+Npred}](#) επτά βουλευτών
 οθηρικών προσλήψεων... {S}ΕΙΤΕ μέσω της [ανάθεσης έργων{ανάθεσης έργων.N+NNG+Nabs}](#) χωρίς διαγωνι
 οθηρικών προσλήψεων... {S}ΕΙΤΕ μέσω της [ανάθεσης έργων{ανάθεσης έργων.N+NNG+Npred}](#) χωρίς διαγων
 οθηρικών προσλήψεων... {S}ΕΙΤΕ μέσω της [ανάθεσης έργων{ανάθεσης έργων.N+NNG+TS}](#) χωρίς διαγωνισμ
 '70 σίγησης.{S} Η Νανά Ησαΐα πέθανε από [ανακοπή καρδιάς{ανακοπή καρδιάς.N+NNG+Nabs}](#), ύστερα από
 '70 σίγησης.{S} Η Νανά Ησαΐα πέθανε από [ανακοπή καρδιάς{ανακοπή καρδιάς.N+NNG+Npred}](#), ύστερα απ
 '70 σίγησης.{S} Η Νανά Ησαΐα πέθανε από [ανακοπή καρδιάς{ανακοπή καρδιάς.N+NNG+TS}](#), ύστερα από μ
 Ακολουθεί η απώλεια των αισθήσεων και η [ανακοπή{ανακοπή.N+NNG+ambigu+Nabs}](#) της καρδιακής λειτου
 Ακολουθεί η απώλεια των αισθήσεων και η [ανακοπή{ανακοπή.N+NNG+ambigu+Npred}](#) της καρδιακής λειτο
 Ακολουθεί η απώλεια των αισθήσεων και η [ανακοπή{ανακοπή.N+NNG+ambigu+TS}](#) της καρδιακής λειτουργγ
 τοπίσουν αυτούς τους μορφασμούς με έναν [ανιχνευτή ψεύδους{ανιχνευτή ψεύδους.N+NNG+Nconc}](#), τον π
 τα εκρηκτικούς μηχανισμούς. {S}Κατά τις [ανταλλαγές πυρών{ανταλλαγές πυρών.N+NNG+Nabs}](#), σκοτώθηκ
 τα εκρηκτικούς μηχανισμούς. {S}Κατά τις [ανταλλαγές πυρών{ανταλλαγές πυρών.N+NNG+Npred}](#), σκοτώθη
 ες προχωρά το υπουργείο Άμυνας για τους [αντιρρησίες συνείδησης{αντιρρησίες συνείδησης.N+NNG+Nhum}](#)
 ά την επίσκεψη εισαγγελέα σταμάτησε την [απεργία πείνας{απεργία πείνας.N+NNG+Nabs}](#)</H3> Σταμάτησ
 ά την επίσκεψη εισαγγελέα σταμάτησε την [απεργία πείνας{απεργία πείνας.N+NNG+Npred}](#)</H3> Σταμάτη
 σιασθεί και στην περίπτωση που υπάρξουν [βεβαιώσεις αποδοχών{βεβαιώσεις αποδοχών.N+NNG+Nabs}](#) σε
 σιασθεί και στην περίπτωση που υπάρξουν [βεβαιώσεις αποδοχών{βεβαιώσεις αποδοχών.N+NNG+Npred}](#) σε
 ία θεάματος, προσκομίζοντας οι ίδιοι το [βιβλιάριο ασθενείας{βιβλιάριο ασθενείας.N+NNG+Nconc}](#) ή
 μηθεί 1.809.862 αυτοκόλλητα για θεώρηση [βιβλιαρίων υγείας{βιβλιαρίων υγείας.N+NNG+Nconc}](#). * Το
 ών, των στρατιωτικών και των ανδρών των [Σωμάτων Ασφαλείας{Σωμάτων Ασφαλείας.N+NNG+Nhumcoll}](#). {S
 εξής: 1.{S} Εκδίδονται ειδικοί Ετήσιοι [Τίτλοι Αποταμίευσης{Τίτλοι Αποταμίευσης.N+NNG+Nconc}](#) (Ε
 ρνητικής προπαγάνδας, την τηλεόραση της [κλειδαρότρυπας{κλειδαρότρυπας.N+NNG+soude+Nconc}](#) που ευ
 ρα ξεφαντώνει στις πίστες καίγοντας σαν [μαρουλόφυλλα{μαρουλόφυλλα.N+NNG+soude+Nconc}](#) τα πεντοχί
 χώρο των φυλακών, να αντιμετωπιστούν τα [στερητικά σύνδρομα{στερητικά σύνδρομα.N+NNG+adjectivation+Nabs}](#)
 χώρο των φυλακών, να αντιμετωπιστούν τα [στερητικά σύνδρομα{στερητικά σύνδρομα.N+NNG+adjectivation+Npred}](#)
 ς του δικαιοκτικού πολιτισμού μας, όπως το [τεκμήριο αθωότητας{τεκμήριο αθωότητας.N+NNG+Nabs}](#), η αν
 θαριστή.{S} Τον ερασιτέχνη Πανιώνιο και [τιμή εκκίνησης{τιμή εκκίνησης.N+NNG+Nconc}](#) που δεν είχε
 0.000 φοιτητές που σπουδάζουν εκτός του [τόπου διαμονής{τόπου διαμονής.N+NNG+Nconc}](#) τους βιώνουν
 εποχή που έχουν βγει ακόμα και έγχρωμοι [φακοί επαφής{φακοί επαφής.N+NNG+Nconc}](#), εξακολουθεί να
 ς Φλωρίνης </H3> Για 6 πίτες * 1 πακέτο [φύλλο κρούστας{φύλλο κρούστας.N+NNG+Nconc}](#), σε θερμοκρα
 πως «δεν πρόκειται να χρησιμεύσουμε ως [φύλλο συκής{φύλλο συκής.N+NNG+Nabs}](#) για τις αποτυχημένε
 πως «δεν πρόκειται να χρησιμεύσουμε ως [φύλλο συκής{φύλλο συκής.N+NNG+Nconc}](#) για τις αποτυχημέν
 πιά.{S} Στραγγίζουμε και στεγνώνουμε με [χαρτί κουζίνας{χαρτί κουζίνας.N+NNG+Nconc}](#). {S}2.{S} Αν
 άνετε άσκηση και μην καπνίζετε ή κάνετε [χρήση ναρκωτικών{χρήση ναρκωτικών.N+NNG+Nabs}](#).{S} Να εί
 άνετε άσκηση και μην καπνίζετε ή κάνετε [χρήση ναρκωτικών{χρήση ναρκωτικών.N+NNG+Npred}](#).{S} Να ε
 κοι της πρωτεύουσας χθες και μάλιστα σε [ώρα αιχμής{ώρα αιχμής.N+NNG+Nabs}](#) από τις 10 έως τις 11
 της Μεταμόρφωσης ιδιαίτερα τις πρωινές [ώρες αιχμής{ώρες αιχμής.N+NNG+Nabs}](#).{S} Οι άλλες τρεις

Les résultats sont, au premier abord, assez satisfaisants, notamment en ce qui concerne la reconnaissance des variantes soudées (e.g. *μαρουλόφυλλα*/feuilles de salade) et la reconnaissance des variantes de type *AN* (e.g. *στερητικά σύνδρομα*/syndromes de sevrage).

La comparaison quantitative des résultats obtenus après l'application du dictionnaire électronique morphologique et des graphes générés au même corpus montre que les graphes générés permettent la reconnaissance d'un nombre plus élevé de formes. Avec le DELACF nous avons obtenu 837 matches, qui correspondent à 2 925 recognized units, tandis qu'avec les graphes générés nous avons obtenu 1 449 matches, qui correspondent à 2 951 recognized units. L'application des graphes générés nous permet en plus d'obtenir des informations syntaxico-sémantiques sur les *N (E+DET:G) N:G* reconnus (e.g. TS, Nconc, Nabs, Npred).

Conclusion

Le principal objectif de cette étude était l'élaboration manuelle de ressources linguistiques pour les $N(E + DET:G)N:G$ du grec moderne en vue de leur reconnaissance lexicale automatique dans les textes écrits. Le cadre théorique que nous avons adopté pour cette recherche nous a été fourni par les travaux de G. Gross (1996a) et de M. Gross (1975, 1977), reposant sur les principes de la grammaire transformationnelle harisienne (Z. S. Harris 1976).

Lors de la première étape de notre recherche, nous avons essayé de définir la notion fondamentale qui régit cette étude : celle de figement. Dans ce but, nous nous sommes servie des critères d'ordre morphologique, syntaxique et sémantique proposés par G. Gross (1996a : 9-22), que nous avons appliqués aux $N(E + DET:G)N:G$ grecs.

Ensuite, nous avons procédé au recensement aussi exhaustif que possible des $N(E + DET:G)N:G$ à partir de sources éditées et numériques. L'étape du recensement nous a permis de rendre compte de la grande hétérogénéité à l'intérieur des $N(E + DET:G)N:G$. Nous avons identifié cinq sous-catégories, à savoir : les noms composés entièrement figés, les syntagmes nominaux semi-figés (ou productifs), les entités nommées, les termes de spécialité et les syntagmes nominaux figés qui font partie d'une phrase simple figée mais qui ont un emploi autonome en dehors de l'expression figée. Le but de cette proposition est de mieux organiser les ressources électroniques des $N(E + DET:G)N:G$ et d'en faciliter la maintenance et la mise à jour. Une fois les $N(E + DET:G)N:G$ recensés, nous avons étudié leur structure lexicale interne et notamment les phénomènes de variation. Les $N(E + DET:G)N:G$ présentent un nombre considérable de variantes. Les variantes peuvent être d'ordre lexical (notamment, à cause de la substitution assez fréquente de formes *savantes* par des formes *démotiques* ou à cause de la substitution de formes grecques par des emprunts), syntaxique (dues à la réduction d'un ou de plusieurs constituants, à l'adjectivation du complément de nom au génitif, à la possibilité de substitution par des formes soudées etc.) ou encore orthographique et phonologique. Nous avons effectué la mise en relation entre les formes variées à l'aide des tables du lexique-grammaire. Nous envisageons d'introduire les variantes dans le dictionnaire électronique morphologique lors d'une étude ultérieure.

L'étude syntaxico-sémantique des $N(E + DET:G)N:G$ a constitué l'étape suivante de notre travail. Nous avons d'abord étudié le rôle du génitif au sein des $N(E + DET:G)N:G$ et la formation du complément de nom sur la base des dérivations syntaxiques régulières et reproductibles. Ensuite, nous avons effectué la distinction entre $N(E + DET:G)N:G$ prédicatifs et $N(E + DET:G)N:G$ non-prédicatifs sur la base de critères formels. Nous avons

proposé la description des $N(E + DET:G)N:G$ prédictifs à l'aide des tables du lexique-grammaire au même titre que les noms prédictifs simples. Enfin, nous avons codé des traits sémantiques élémentaires pour les $N(E + DET:G)N:G$ de cette étude. Nous considérons que la description des propriétés sémantiques est importante, car elle permettrait à un outil d'analyse utilisant le lexique-grammaire d'avoir accès à un lexique structuré sémantiquement de façon cohérente avec la description des arguments dans les tables du lexique-grammaire.

Au niveau du traitement automatique en vue de la reconnaissance lexicale automatique des $N(E + DET:G)N:G$ grecs dans les textes écrits, nous avons élaboré des dictionnaires électroniques morphologiques pour les $N(E + DET:G)N:G$ binaires et les $N1(E + DET:G)A:G N2:G$ complexes (ou surcomposés). Nous avons proposé la solution des dictionnaires électroniques morphologiques pour les noms composés entièrement figés, les termes de spécialité et les syntagmes nominaux figés qui font partie d'une phrase simple figée mais qui ont un emploi autonome en dehors de l'expression figée. Nous avons effectué la représentation formalisée des variantes syntaxiques et le codage sémantique de ces mêmes sous-catégories à l'aide des tables du lexique-grammaire, que nous avons par la suite converties automatiquement en graphe d'automates finis à l'aide d'Unitex. Nous avons construit des graphes lexicalisés pour les syntagmes nominaux semi-figés (ou productifs). Enfin, concernant les entités nommées, nous avons adopté une solution provisoire : celle de la constitution d'une liste séparée, qui pourrait être exploitée lors d'une étape ultérieure.

Nous envisageons de compléter cette recherche par une étude complète des noms composés complexes (ou surcomposés) de toutes les catégories morphosyntaxiques identifiées. Nous envisageons également de compléter l'élaboration du programme de flexion pour l'ajout automatique de traits sémantiques et la représentation de toutes les variantes dans le dictionnaire électronique morphologique des noms composés. Enfin, nous envisageons d'intégrer les entités nommées au dictionnaire électronique morphologique en leur assignant un code spécifique.

Notre étude apporte une contribution à plusieurs disciplines telles que la linguistique, la lexicographie, la traduction (humaine ou automatique), l'informatique linguistique, l'enseignement du grec moderne en tant que langue étrangère. Au niveau linguistique, notre étude contribue notamment à l'approfondissement des phénomènes de composition nominale et de figement.

Les noms composés, faisant partie des expressions figées d'une langue, ont toujours posé le problème d'enregistrement dans les dictionnaires édités. Où doit-on, par exemple, enregistrer le $N(E+DET:G)N:G$ *ανιχνευτής ψεύδους*/détecteur de mensonge ? A l'entrée de *ανιχνευτής*/détecteur ou à celle de *ψεύδος*/mensonge ? Les lexicographes ont du mal à répondre à cette question. C'est pourquoi ces entrées n'ont pas toujours un statut autonome. Elles sont souvent éliminées des dictionnaires édités ou alors elles sont codées de manière incohérente. Il faut signaler aussi que les dictionnaires usuels ne disposent pas d'un répertoire adéquatement chiffré de $N(E+DET:G)N:G$. Le dictionnaire électronique morphologique et les tables du lexique-grammaire que nous avons construites dans le cadre de cette recherche pourraient servir à la mise à jour des dictionnaires usuels.

Nos données linguistiques présentent également un degré de formalisation aisément traitable dans le cadre d'applications automatiques telles que l'indexation et la traduction. Elles pourraient donc être exploitées par des systèmes de recherche documentaire, d'extraction d'information, d'analyse automatique des textes, de traduction automatique et dans des correcteurs morpho-syntaxiques.

Enfin, concernant l'enseignement du grec en tant que langue étrangère, les ressources que nous avons élaborées dans le cadre de cette étude pourraient contribuer à l'apprentissage des expressions figées, qui ont toujours constitué un problème majeur lors de l'apprentissage d'une langue étrangère.

Bibliographie

ABEILLE, Anne ; L., CLEMENT ; F. TOUSSENEL. 2003. « Building a French Treebank », In *Treebanks*, A. Abeillé (Ed.), Dordrecht : Kluwer, pp. 165-188.

ALLERTON, D. 1987. « The linguistic and sociolinguistic status of proper names », In *Journal of Pragmatics*, vol. 11 : 61-92.

ANASTASSIADIS-SYMEONIDIS, Anna. 1986. *Η νεολογία στην κοινή νεοελληνική*, Thèse de Doctorat, Thessaloniki : Université Aristote de Thessaloniki.

ANASTASSIADIS-SYMEONIDIS, Anna. 1989. « Le syntème du type adj.+nom en grec moderne », In *Journal of Applied Linguistics* 5, Thessaloniki, pp. 5-14.

ANASTASSIADIS-SYMEONIDIS, Anna. 1994. « Les composés du type AN en grec moderne », *Lingua Franca* 1, Brindisi, pp. 37-56.

ANASTASSIADIS-SYMEONIDIS, Anna. 1996. «Η νεοελληνική σύνθεση», In *Ζητήματα Νεοελληνικής Γλώσσας – Διδακτική Προσέγγιση*, G. Katsimali, F. Kavoukopoulos (éds.), Réthymnon : Université de Crète, pp. 97-120.

ANASTASSIADIS-SYMEONIDIS, Anna. 2003. « Que peut-il arriver à une expression figée », In *Cahiers de Lexicologie* 82 : 1, Paris : Didier Erudition, pp. 1-9.

ANASTASSIADIS-SYMEONIDIS, Anna. 2004. « Le lexique-grammaire du grec moderne », *Syntax, Lexis and Lexicon-Grammar. Papers in honour of Maurice Gross*, *Linguisticae Investigationes Supplementa*, 24, Amsterdam/Philadelphia : Benjamins, pp. 11-22.

ANASTASSIADIS-SYMEONIDIS, Anna ; Asimakis FLIATOURAS. 2003. «Η διάκριση [λόγιο] και [λαϊκό] στην ελληνική γλώσσα. Ορισμός και ταξινόμηση», In *Proceedings of the 6th ICGL, 18-21 September 2003*, Rethymnon.

ANSCOMBRE, Jean-Claude. 1991. « La détermination zéro : quelques propriétés », *Langages* 102, Paris : Larousse, pp. 103-124.

BABINIOTIS, Georgios. 1985. *Εισαγωγή στη σημασιολογία*, Athènes.

BALLY, Charles. 1921. *Traité de stylistique française*, Paris : Klincksieck.

BALLY, Charles. 1965. *Linguistique générale et linguistique française*, Berne : Francke, 4^{ème} éd.

BELANGER, Alain. 1997. *Le problème des groupes N de N et sa solution en Lexique-Grammaire (Les constructions N0 avoir N de N W)*, Thèse de doctorat, Paris : Université de Paris-Nord.

BIKEL, D. M. ; S. MILLER ; R. SCHWARTZ ; R. WEISCHEDEL. 1997. "Nymble: a high-performance learning name-finder", In *Proceedings of the 5th Conference on Applied Natural language processing, 31/03-03/04 1997*, Washington, DC: Morgan Kaufman Publishers Inc.

BLANC, Olivier ; Matthieu, CONSTANT. 2005. "Lexicalization of Grammars with Parameterized Graphs", In *Proceedings of RANLP 2005, 21-23 september 2005*, G. Angelova *et. al* (éds.), Borovets, pp. 117-121.

BLANCO, Xavier ; Dolores, CATALÀ. 1998/1999. « Quelques remarques sur un dictionnaire électronique d'adverbes composés en espagnol », *Lingvisticae Investigationes* 22, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 213-232.

BLANCO, Xavier ; Pierre-André, BUVET. 1999. « A propos de la traduction automatique des déterminants de l'espagnol et du français », In *Méta* 44 : 4, Montréal : Les Presses de l'Université de Montréal, pp. 525-545.

BLANCO, Xavier ; Pierre-André, BUVET ; Zoé, GAVRIILIDOU. 1999. « Analyse comparée des déterminants et des modifieurs figés en espagnol, français, grec. Vers un dictionnaire électronique », *Studies in Greek Linguistics. Proceedings of the 20th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, Thessaloniki, pp. 87-99.

BLANCO, Xavier. 2001. « Regroupements sémantiques dans un dictionnaire d'adverbes composés en espagnol », *Lingvisticae Investigationes* 24 : 2, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 167-182.

BLANCO, Xavier. 2002. « Les déterminants figés », *Langages* 145, Paris : Larousse, pp. 61-82.

BOONS, Jean-Paul. 1971. « Métaphore et baisse de la redondance », *Langue Française* 11, Larousse, Paris, pp. 15-16.

BOONS, Jean-Paul. 1979. "On figurative meaning and redundancy", *SMIL, Journal of Linguistic Calculus* 1-2, Stockholm : Skriptor, pp. 101-104.

BOONS, Jean-Paul ; Alain, GUILLET ; Christian, LECLÈRE. 1976. *La structure des phrases simples en français: constructions intransitives*, Genève : Droz.

BORILLO, Andrée. 1971a. *Les constructions réciproques et symétriques en français*, Thèse de 3^{ème} cycle, Aix : Université d'Aix.

BORILLO, Andrée. 1971b. « Remarques sur les verbes symétriques français », *Langue Française* 11, Paris : Larousse, pp. 17-31.

BRUNOT, Ferdinand ; Charles BRUNEAU. 1949. *Précis de la grammaire historique de la langue française*, Paris : Masson.

BUVET, Pierre-André ; Gaston, GROSS. 1995. « Comparaison et expression du haut degré dans le groupe nominal », *Faits de langues* 5, Paris : PUF, pp. 83-88.

BUVET, Pierre-André ; Lim, JUNG-HAE. 1996. « Les déterminants nominaux aspectuels », *Lingvisticae Investigationes* 20 : 2, Amsterdam/Philadelphia : John Benjamins Publishing Co, pp. 271-285.

BUVET, Pierre-André. 1993. *Les déterminants nominaux quantifieurs*, Thèse de doctorat, Paris : Université Paris XIII.

BUVET, Pierre-André, 1994. « Les déterminants nominaux », *Lingvisticae Investigationes* 18 : 1, Amsterdam/Philadelphia : John Benjamins Publishing Co, pp. 121-150.

BUVET, Pierre-André, 1995. « Les compléments de nom intensifs », *Neophilologica* 12, Katowice : Université de Silésie, pp. 91-102.

BUVET, Pierre-André. 1998. « Détermination et classes d'objets », *Langages* 131, Paris : Larousse, pp. 91-102.

BUVET, Pierre-André. 2001. « Les déterminants intensifs », *Lingvisticae Investigationes Supplementa* 23, *Détermination et Formalisation*, X. Blanco ; P.-A. Buvet ; Z. Gavriilidou (éds.), Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 101-114.

CABRE, Maria Teresa. 1998. *La terminologie, théorie, méthodes et applications*, Presses Universitaires d'Ottawa.

CATACH, Nina. 1981. *Orthographe et lexicographie. Les mots composés*, Paris : Nathan.

CHINCHOR, N. (1998), « MUC-7 Named Entity Task Definition (version 3.5) », *Proceedings of the 7th Message Understanding Conference (MUC-7)*, 19 Avril -1er Mai 1998, Fairfax, VA.

CHOMSKY, Noam. 1965. *Aspects of the Theory of Syntax*, Cambridge : MIT Press, 102 p.

CHOMSKY, Noam. 1981. *Lectures on Government and Binding. The Pisa Lectures*, 2. Dordrecht : Foris.

CLAIRIS, Christos ; Georgios, BABINIOTIS. 1996. *Γραμματική της Νέας Ελληνικής. Δομολειτουργική-Επικοινωνιακή : Τόμος Α΄. Το όνομα. Αναφορά στον κόσμο της πραγματικότητας*, Athènes : Ellinika Grammata.

CLAIRIS, Christos ; Georgios, BABINIOTIS. 1999. *Γραμματική της Νέας Ελληνικής. Δομολειτουργική-Επικοινωνιακή : Τόμος Β΄. Το ρήμα. Η οργάνωση του μηνύματος*, Athènes : Ellinika Grammata.

CLAIRIS, Christos ; Georgios, BABINIOTIS. 2001. *Γραμματική της Νέας Ελληνικής. Δομολειτουργική-Επικοινωνιακή : Τόμος Γ΄. Τα επιρρηματικά στοιχεία. Η εξιδείκευση του μηνύματος*, Athènes : Ellinika Grammata.

CONSTANT, Matthieu. 2003. *Grammaires locales pour l'analyse automatique de textes : Méthodes de construction et outils de gestion*, Thèse de doctorat, Paris : Université de Marne-la-Vallée.

CONSTANT, Matthieu ; Anastasia, YANNAKOPOULOU. 2003. « Le dictionnaire électronique du grec moderne : conception et développement d'outils pour son enrichissement et sa validation », In *Studies of Greek Linguistics, Proceedings of the 23rd Annual Meeting of the Department of Linguistics*, Faculty of Philosophy, Aristotle University of Thessaloniki, Vol II, pp. 783-791.

CONSTANT, Matthieu. 2007. « GraalWeb ou accéder à une bibliothèque décentralisée de grammaires locales », In *Actes du Colloque Bases de données lexicales : construction et applications*, Montréal, pp. 79-87.

CORBLIN, Francis. 1983. « Définis et démonstratifs dans la reprise immédiate », *Le français moderne* 51 : 2, CILF, pp. 118-133.

CORBLIN, Francis. 1987. *Indéfini, défini et démonstratif*, Genève : Droz.

CORBLIN, Francis. 1989. « Spécifique-générique : un modèle pour les indéfinis », *Modèles Linguistiques* 11 : 2, Lille : Presses Universitaires, pp. 11-35.

COURTOIS, Blandine. 1990. « Un système de dictionnaires électroniques pour les mots simples du français », *Langue française* 87, Paris : Larousse, pp. 11-22.

COURTOIS, Blandine. 1994. *Marques lexicales, syntaxiques, sémantiques dans le DELAS V08/E*.

COURTOIS, Blandine. 1996. « Les formes ambiguës de la langue française », *Linguisticæ Investigationes* 20 : 1, Amsterdam/Philadelphia : John Benjamins Publishing and Co, pp. 167-202.

DAILLE, B. 1995. *Repérage et extraction de terminologie par une approche mixte statistique et linguistique*, TAL, ATALA, 36(1-2) : 101-118.

DAILLE, B. 1996. « Study and Implementation of Combined Techniques for Automatic Extraction of Terminology », in P. Resnik and J. Klavans (eds) *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, MIT Press, p. 46-66.

DAILLE, B. ; E. MORIN. 2000. « Reconnaissance automatique des noms propres de la langue écrite : les récentes réalisations », In Maurel D. et Gueunthner F. (éds), *Traitement Automatique des Langues*, vol. 41/3, pp. 601-621.

DANLOS, Laurence. 1980. *Représentation d'informations linguistiques : Constructions N être Prép X*, Thèse de 3^{ème} cycle, Paris : Université Paris VII, Paris : LADL.

DANLOS, Laurence. 1981. « La morphosyntaxe des expressions figées », *Langages* 63, *Formes syntaxiques et prédicats sémantiques*, A. Guillet ; C. Leclère (éds.), Paris : Larousse, pp. 53-74.

DANLOS, Laurence. 1986. *Génération automatique de textes en langues naturelles*, Paris : Masson.

DANLOS, Laurence. 1988. « Les expressions figées construites avec le verbe support être Prép », *Langages* 90, Paris : Larousse, pp. 23-28.

DARMESTETER, Arsène. 1874. *Traité de la formation des mots composés*, Paris : Bouillon.

DELVEROUDI, Rhéa. 1995. « Ο συμπληρωματικός δείκτης και », *Studies in Greek Linguistics, Proceedings of the 15th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, Thessaloniki, pp. 281-291.

DUGAS, André. 1973. « Les compléments d'instrument », *Cahier de Linguistique 2*, Problèmes de sémantique, Montréal : Service des publications de l'Université du Québec à Montréal, pp. 168-180.

DUGAS, André; Anna Maria, DI SCIULLO. 1985. « Le rôle des déterminants dans les expressions figées de langues romanes », *Actes du colloque international "La Locution", Le Moyen Français 14-15*, Montréal : McGill, pp. 56-70.

FIALA, P.; B. HABERT. 1989. « La langue de bois en éclat : les défigements dans les titres de la presse quotidienne française », In *MOTS 21*, Presses de la Fondation Nationale des Sciences Politiques, pp. 83-99.

FIRENZE, Anna ; Béatrice, PELLETIER. 1989. « Recherche d'une description syntaxique contrastive des noms composés N de N du français et N di N, N da N de l'italien », In *Actes du colloque « La description des langues naturelles en vue d'application informatiques »*, Québec : Université Laval.

FLAUX, Nelly ; Danièle, VAN DE VELDE. 2000. *Les noms en français : esquisse de classement*, Paris : Ophrys.

FOTOPOULOU, Aggeliki. 1990. « Classification des phrases figées ayant un complément libre. Problèmes liés aux compléments datifs », In *Studies in Greek Linguistics, Proceedings of the 11th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, Thessaloniki.

FOTOPOULOU, Aggeliki. 1993a. *Une classification des phrases à compléments figés en grec moderne. Etude morphosyntaxique des phrases figées*, Thèse de Doctorat, Paris : Université Paris VIII.

FOTOPOULOU, Aggeliki. 1993b. «Ανάλυση των συστατικών των στερεότυπων προτάσεων. Παρατηρήσεις σχετικά με την ταξινόμησή τους», In *Studies in Greek Linguistics, Proceedings of the 14th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki, 27-29 april 1993*, Thessaloniki, pp. 276-292.

FOTOPOULOU, Aggeliki. 1993c. « Traitement du cas génitif dans une classification des phrases à compléments figés du grec moderne », In *Lingvisticae Investigationes* 17 : 2, Amsterdam-Philadelphia : John Benjamins, pp. 259-279.

FOUROUR, N. 2002. « Nemesis, un système de reconnaissance incrémentielle des entités nommées pour le français », In *Actes de la 9^{ème} Conférence Nationale sur le Traitement Automatique des Langues Naturelles (TALN 2001)*, Nancy, vol. 1, pp. 265-274.

FRIBURGER, N. 2002. *Reconnaissance automatique des noms propres : Application à la classification automatique des textes journalistiques*, Thèse de doctorat, Université de Tours.

GAATONE, David. 1997. « La locution: Analyse interne et analyse globale », *La locution entre langue et usages*, M. Martins-Baltar (éd.), Paris : ENS éditions, Fontenay/Saint-Cloud, pp. 165-177.

GAATONE, David. 2000. « A quoi sert la notion d'« expression figée » ? », In *BULAG. Lexique, Syntaxe et Sémantique. Mélanges offerts à Gaston Gross à l'occasion de son soixantième anniversaire*, A.-P. Buvet et al. (éds), Besançon.

GARRIGUES, Mylène. 1997. « Une méthode de désambiguïsation locale nom/adjectif pour l'analyse automatique de textes », *Langages* 126, Paris : Larousse, pp. 60-78.

GARRIGUES, Mylène. 1998. « La place de l'adjectif en français et sa formalisation dans un dictionnaire électronique », *Travaux de linguistique* 37, B. Lamiroy (éd.), Paris : Duculot.

GAVRIILIDOU, Zoé. 1997. *Etude comparée des suites NN en français et en grec (Elaboration d'un lexique bilingue)*. Thèse de doctorat, Paris : Université Paris XIII.

GAVRIILIDOU, Zoé. 1998. « Les déterminants nominaux figés en grec moderne : Une première approche », In *Actes des Premières Rencontres Linguistiques Méditerranéennes sur le figement lexical*, S. Mejri ; G. Gross ; A. Clas ; T. Baccouche (éds.), Tunis : Publications de la faculté des lettres de la Manouba, pp. 35-50.

GAVRIILIDOU, Zoé. 2001. « Structures *Dét N1 N2* et détermination figée », *Lingvisticae Investigationes Supplementa* 23, *Détermination et Formalisation*, X. Blanco ; P.-A. Buvet ; Z. Gavriilidou (éds.), Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 163-175.

GAVRIILIDOU, Zoé. 2002. « La détermination des noms de sentiment en grec moderne », *Langages* 145, Paris : Larousse, pp. 83-96.

- GIRY-SCHNEIDER, Jacqueline. 1978. *Les nominalisations en français*, Genève : Droz.
- GIRY-SCHNEIDER, Jacqueline. 1987. *Les prédicats nominaux en français. Les phrases simples à verbe support*, Genève : Droz.
- GOUADEC, Daniel. 1990. *Terminologie-Constitution des données*, Paris : AFNOR Gestion.
- GOUTSOS, Dionyssis. 2002. « L'utilisation des corpus textuels électroniques lors de l'analyse du discours », *Recherches en Linguistique Grecque II : Actes du 5ème Colloque International de Linguistique Grecque, Sorbonne, (13-15 septembre 2001)*, Paris : L'Harmattan, pp. 219-222.
- GRASS, Thierry. 2000. « Les noms propres d'association et d'organisation : traduction et traitement automatique », In *Les Nouveaux Cahiers d'Allemand* n°2/2008, Nancy, pp. 161-174.
- GRASS, Thierry ; Denis, MAUREL. 2008. « Typologie et traductibilité des noms propres de l'allemand vers le français à partir d'un corpus journalistique », In Maurel D. et Gueunthner F. (éds), *Traitement Automatique des Langues*, vol. 41/3, pp. 643-669.
- GRESILLON, A. ; D. MAINGUENEAU. 1984. « Polyphonie, proverbe et détournement », *Langages* 73, Paris : Larousse, pp. 112-125.
- GREVISSE, Maurice. 1986. *Le bon usage*, (12^{ème} éd. refondue par A. Goose), Paris-Gembloux : Duculot.
- GROSS, Gaston. 1981. « Les prépositions composées », *Actes du troisième colloque franco-allemand de linguistique théorique « Analyse des prépositions »*, Christoph Schwarze (éd.), Tübingen : Max Niemeyer Verlag, pp. 29-39.
- GROSS, Gaston. 1986a. « Étude syntaxique du déterminant possessif », *Recherches Linguistiques* 11, Paris : Klincksieck.
- GROSS, Gaston. 1986b. *Typologie des noms composés*, Rapport A.T.P. du CNRS.
- GROSS, Gaston. 1988. « Réflexions sur la notion de locution conjonctive », *Langue Française* 77, *Syntaxe des connecteurs*, G. Gross ; M. Piot (éds.), Paris : Larousse.
- GROSS, Gaston. 1990. « Définition des noms composés dans un lexique-grammaire », *Langue Française* 87, Paris : Larousse.

GROSS, Gaston. 1991. « Analyse des groupes N de N », In *Actes du colloque « Les industries de la langue »*, Nantes.

GROSS, Gaston ; André, Dugas. 1991. « Analyse automatique des suites N de N », In *Actes du colloque « Informatique et Langue Naturelle »*, Nantes, pp. 391-415.

GROSS, Gaston. 1994. « Classes d'objets et description des verbes », *Langages* 115, Paris : Larousse, pp. 15-31.

GROSS, Gaston. 1995a, « Pour une typologie de la détermination nominale », *Tendances récentes en linguistique française et générale, volume dédié à David Gaatone, Lingvisticae Investigationes Supplementa* 20, H. Bat-Zeev Shyldkrot ; L. Kupferman (éds.), Amsterdam/Philadelphia : John Benjamins Publishing Co, pp. 161-171.

GROSS, Gaston. 1995b, « A propos de la notion d'humain », *Lingvisticae Investigationes Supplementa* 17, Amsterdam/Philadelphia : John Benjamins, p. 71-80.

GROSS, Gaston. 1996a. *Les expressions figées en français : noms composés et autres locutions*, Paris : Ophrys.

GROSS, Gaston. 1996b. « Prédicats nominaux et compatibilité aspectuelle ». *Langages* 121, Paris : Larousse, pp. 54-72.

GROSS, Gaston. 1997. « Du bon usage de la notion de locution », *La locution entre langue et usages*, M. Martins-Baltar (éd.), Paris : ENS éditions, Fontenay/Saint-Cloud, pp. 201-255.

GROSS, Gaston. 1999a. « Etude lexicale des locutions conjonctives », *Mémoires de la Société de Linguistique de Paris* 7 (Nouvelle Série), pp. 23-40.

GROSS, Gaston. 1999b. « Sémantique lexicale et connecteurs », *Langages* 136, Paris : Larousse, pp. 76-84.

GROSS, Gaston. 2002. « Locutions conjonctives et détermination », *Langages* 145, Paris : Larousse, pp. 9-20.

GROSS, Maurice. 1964. « Sur certains procédés de définition de langages formels », In *Automata Theory*, E.R. Caianiello (éd.), New York : Academic Press, pp. 181-200.

GROSS, Maurice ; André, LENTIN. 1967. *Notions sur les grammaires formelles*, Paris : Gauthier-Villars.

- GROSS, Maurice. 1967. « Sur une règle de cacophonie », *Langages* 7, Paris : Larousse.
- GROSS, Maurice. 1968. *Grammaire transformationnelle du français : 1. Syntaxe du verbe*, Paris : Cantilène.
- GROSS, Maurice. 1969. « Remarques sur la notion d'objet direct en français », *Langue Française* 1, Paris : Larousse, pp. 63-73.
- GROSS, Maurice. 1971. « Grammaire transformationnelle et enseignement du français », *Langue Française* 11, Paris : Larousse, pp. 4-14.
- GROSS, Maurice. 1975. *Méthodes en syntaxe*, Paris : Hermann.
- GROSS, Maurice. 1977. *Grammaire transformationnelle du français : 2. Syntaxe du nom*, Paris : Larousse, Réimpression 1986 : Cantilène.
- GROSS, Maurice. 1981. « Les bases empiriques de la notion de prédicat sémantique », *Langages* 63, *Formes syntaxiques et prédicats sémantiques*, A. Guillet ; C. Leclère (éds.), Paris : Larousse, pp. 7-52.
- GROSS, Maurice. 1982. « Une classification des phrases figées du français », *Revue Québécoise de Linguistique* 11 : 2, Montréal : Presses de l'Université du Québec à Montréal, pp. 151-185.
- GROSS, Maurice. 1983. « Sur quelques types de coréférence contrainte », *Revue Romane* 24, *Analyses grammaticales du français, Etudes publiées à l'occasion du 50ème anniversaire de Carl Vikner*, pp. 116-129.
- GROSS, Maurice. 1985. « Sur les déterminants dans les expressions figées », *Langages* 79, *Générique et généricité*, M. Calmiche ; G. Kleiber (éds), Paris : Larousse, pp. 89-117.
- GROSS, Maurice. 1988a. « La phrase élémentaire et ses composants. Une discussion de quelques exemples », In *Travaux de linguistique* 17, *La prédication seconde*, Paris/Gembloux : Duculot, pp. 13-32.
- GROSS, Maurice. 1988b. « Sur les phrases figées complexes du français », *Langue Française* 77, Paris : Larousse, pp. 47-70.
- GROSS, Maurice. 1988c. « Les limites de la phrase figée », *Langages* 90, Paris : Larousse, pp. 7-22.

GROSS, Maurice. 1989. *Grammaire transformationnelle du français : 4. Syntaxe des phrases figées*, Paris : Canitlène.

GROSS, Maurice. 1990. *Grammaire transformationnelle du français : 3. Syntaxe de l'adverbe*, Paris : ASSTRIL.

GROSS, Maurice. 1995. « Une grammaire locale de l'expression des sentiments », *Langue Française* 105, Paris : Larousse, pp. 70-87.

GROSS, Maurice. 1995. « Quelques considérations sur les marques sémantiques d'un dictionnaire électronique », In *Lexiques-Grammaires comparés et traitements automatiques – Actes du deuxième colloque international LGC*, Montréal, pp. 9-29.

GROSS, Maurice. 1996. « Les formes être Prép X du français », *Linguisticae Investigationes* 20 : 2, Amsterdam-Philadelphia : John Benjamins, pp. 217-270.

GROSS, Maurice. 1997a. « Synonymie, morphologie dérivationnelle et transformations », *Langages* 128, Paris : Larousse, pp. 72-90.

GROSS, Maurice. 1997b. « The Construction of Local Grammars », In *Finite State Language Processing*, E. Roche ; Y.Schabes (éds.), Cambridge, Mass. : The MIT Press, pp. 329-352.

GROSS, Maurice. 2001a. « Les ambiguïtés », *Linguisticae Investigationes* 24 : 1, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 3-41.

GROSS, Maurice. 2001b. « Grammaires locales de déterminants nominaux », *Linguisticae Investigationes Supplementa* 23, *Détermination et Formalisation*, X. Blanco ; P.-A. Buvet ; Z. Gavriilidou (éds.), Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 177-194.

GROSS, Maurice. 2002. « Les déterminants numéraux, un exemple : les dates horaires », *Langages* 145, Paris : Larousse, pp. 21-37.

GUILBERT, Louis. 1971. « De la formation des unités lexicales », *Grand Larousse de la langue française*, t. 1, Paris : Larousse, pp. IX-LXXXI

GUILLET, Alain ; Christian, LECLÈRE. 1981. « Restructuration du groupe nominal », *Langages* 63, Paris : Larousse, pp. 99-125.

GUILLET, Alain ; Christian, LECLÈRE. 1992. *La structure des phrases simples en français. 2 : les constructions transitives locatives*, Genève : Droz.

GUIRAUD, P. 1980. *Les locutions françaises*, Paris : PUF : *Que sais-je* 903.

HARRIS, Zellig S. 1964. *Elementary Transformations*, Philadelphie : University of Pennsylvania, *TDAP* 54, réimprimé dans *Papers in Structural and Transformational Linguistics*, 1970, Reidel : Dordrecht.

HARRIS, Zellig S. 1968. *Mathematical Structures of Language*, New York : Wiley, C, FUCHS (trd.), *Structures mathématiques du langage*, Paris : Dunod.

HARRIS, Zellig S. 1976. *Notes du cours de syntaxe*, Paris : Editions du Seuil.

HJEMSLEV, L. 1935-37. « La catégorie des cas », *Acta jutlandica* VII, 1, P; 1-184, IX p. 1-78.

HOBBS, J. ; D. APPELT ; J. BEAR ; D. ISRAEL ; M. KAMEYAMA ; M. STICKEL ; M. TYSON. 1996. « FASTUS : a cascaded finite-state transducer for extracting information from natural-language text », In Roche E. et Schabes Y. (éds), *Finite State Devices for Natural Language Processing*, Cambridge, USA : MIT Press, pp. 383-406.

HOLTON, David ; Peter, MACKRIDGE ; Ειρήνη, Φιλippάκη-WARBURTON. 2000. *Γραμματική της Ελληνικής Γλώσσας*, V. Spyropoulos (trd.), Athènes : Editions Patakis.

IOANNIDOU, Kyriaki ; Ionas, MICHAILIDIS ; Periklis POLITIS ; Stavroula VOYATZI. 2005. « Paramétrage du corpus grec numérique par genre textuel : le discours journalistique », *25^e colloque international sur le lexique et la grammaire*, Liverpool (communication orale)

IORDANIDOU, Anna. ; Irène, TSAMADOU. 1985. « Κοινωνιογλωσσολογικοί προβληματισμοί πάνω στη μέθοδο συλλογής του γλωσσικού υλικού (corpus) », *Studies in Greek Linguistics, Proceedings of the 6th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, Thessaloniki, pp. 297-306.

JACKENDOFF, F. Ray. 1972. *Semantic Interpretation in Generative Grammar*, Cambridge : MIT Press.

JAKOBSON, Roman. 1936. « Beitrag zur allgemeinen Kasuslehre, Gesamtbedeutungen der rüssischen Kasus », In *Travaux du Cercle Linguistique de Prague*, 6.

JAKOBSON, Roman. 1971. « Two aspects of language and two types of aphasic disturbances », *Selected Writings* 2, Berlin : Mouton de Gruyter, pp. 239-259.

KALAMPOKIS, Alkis ; Tita, KYRIACOPOULOU. 2005. « L'évolution de la langue grecque : Etude statistique à partir des corpus », *24th International Conference « Lexis & Grammar »*, 15-18 septembre 2005, Liverpool.

KAYNE, Richard. 1975. *French Syntax*, Cambridge, Mass. : MIT Press (trd. en français 1977: *Syntaxe du français. Le cycle transformationnel*, Paris : Le Seuil).

KLEEN, S.-C. 1956. « Representation of events in nerve nets and finite automata », In *Automata Studies*, C. E. Shannon ; J. McCarthy (éds.), Princeton : University Press of Princeton.

KLEIBER, Georges. 1985. « La généricité autrement : le traitement ontologique de GN », *Travaux de linguistique et de littérature* 23, pp. 307-331.

KLEIBER, Georges. 1988. « Généricité et raisonnement par défaut », *Le français moderne* 56 : 1/2, pp. 1-16.

KLEIBER, Georges. 1989. « Généricité et typicalité », *Le français moderne* 57 : 3/4, Paris : CILF, pp. 127-154.

KLEIBER, Georges. 1990. *L'article LE générique. La généricité sur le mode massif*, Genève-Paris : Droz.

KLEIBER, Georges. 1995. « Sur les (in)définis en général et les SN (in)définis en particulier », *Bulletin de la Société Linguistique de Paris* 90 : 1, pp. 21-51.

KLEIBER, Georges. 1998. « Quand le tout est de la partie », *Mots chiffrés et déchiffrés*, Mellet, S ; M. Vuillaume (éds.), Paris : Honoré Champion Editeur, pp. 549-565.

KLEIBER, Georges. 2001. « Déterminants indéfinis ou quand les faibles jouent aux forts », *Linguisticae Investigationes Supplementa* 23, *Détermination et Formalisation*, Xavier, B. ; P.-A. Buvet ; Z. Gavriilidou (éds.), Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 195-218.

KLEIN, Jean ; Jacques, LABELLE ; Béatrice, LAMIROY ; Christian, LECLERE. 2003. « Les expressions verbales figées dans quatre variétés de français : le projet BFQS », *Cahiers de lexicologie*, 83(2), pp. 153-172.

KLEIN, Jean ; Béatrice, LAMIROY. 2005. « Le problème central du figement est le semi-figement ». *Linx*, (53), pp. 135-154.

Lamiroy, B., Klein, J. (2005). Le problème central du figement est le semi-figement. *Linx*, (53), 135-154.

KLEIN, Jean-René ; Corine, ROSSARI. 2003. « Figement et variations en français de Belgique, de France, du Québec et de Suisse », *Lingvisticæ Investigationes* 26 : 2, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 203-214.

KYRIACOPOULOU, Tita. 1989. « Le lexique-grammaire : les verbes de mouvement et les verbes de communication », *Proceedings of the 9th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki, (18-20 april 1988)*, Thessaloniki. pp. 427-441.

KYRIACOPOULOU, Tita. 1990. *Les dictionnaires électroniques. La flexion verbale en grec moderne*, Thèse de Doctorat, Paris : Université Paris VIII.

KYRIACOPOULOU, Tita ; Safia, MRABTI ; Anastasia, YANNACOPOULOU. 2002. « Le dictionnaire électronique des noms composés en grec moderne », *Lingvisticæ Investigationes* 25 : 1, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 7-28.

KYRIACOPOULOU, Panayota-Tita. 2003. *Analyse automatique des textes écrits : le cas du grec moderne*, Habilitation à diriger des recherches, Université de Marne-la-Vallée.

KYRIAKOPOULOU, Anthie. 2008. « Méthodes de recensement des noms composés $N(E+DET:G)N:G$ en grec moderne ». *Actes du 27^{ème} Colloque International sur le Lexique et la Grammaire*, 10-13 septembre 2008, L'Aquila.

LABELLE, Jacques. 1986. « Grammaire des noms de maladies », *Langue française* 69, Paris : Larousse.

LAMIROY, Béatrice. 2003. « Les notions linguistiques de figement et de contrainte », *Lingvisticæ Investigationes* 26 : 1, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 1-14.

LAMIROY, Béatrice. 2006. « Le français de Belgique et les locutions verbales figées », *Revue belge de philologie et d'histoire*, 84(3), pp. 829-844.

LAPORTE, Eric. 1992. « Adjectifs en -ant dérivés de verbes », *Langue Française* 96, Paris ; Larousse.

LAPORTE, Eric. 1988. « La reconnaissance des expressions figées lors de l'analyse automatique », *Langages* 90, pp. 117-126.

LAPORTE, Eric. 1995. « Levée d'ambiguïtés par grammaires locales », *Linguisticae Investigationes Supplementa* 17, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 97-114.

LAPORTE, Eric ; Anne, MONCEAUX. 1999. « Elimination of lexical ambiguities by grammars : the ELAG system », *Linguisticae Investigationes* 22, C. Fairon (éd.), Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 341-367.

LAPORTE, Eric. 2001. « Reduction of lexical ambiguity », *Linguisticae Investigationes* 24: 1, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 67-103.

LAPORTE, Eric ; Takuya, NAKAMURA ; Stavroula, VOYATZI. 2008. "A French Corpus Annotated for Multiword Nouns", In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08), Towards a Shared Task for Multiword Expressions (MWE 2008)*, Marrakech, Maroc, pp. 27-30.

LECLÈRE, Christian. 1990. « Organisation du lexique-grammaire des verbes français », *Langue Française* 87, Paris : Larousse, pp. 112-122.

LECLÈRE, Christian. ; Carlos, SUBIRATS-RÜGGERBERG. 1991. « A Bibliography of Studies on Lexicon-Grammar », *Linguisticae Investigationes* 15 : 2, Amsterdam/Philadelphia : John Benjamins, pp. 345-409.

LECLÈRE, Christian. 2002. « Emplois verbaux, distributions, métaphores », *Langue Française* 134, Paris : Larousse, pp. 78-89.

LE MEUR, C. ; S. GALLIANO ; E. GEOFFROIS. 2004. « Conventions d'annotations en Entités Nommées », *ESTER*, <http://www.afcp-parole.org/ester/publis.html>, pp. 6-10.

LE PESANT, Denis ; Michel, MATHIEU-COLAS. 1998. « Introduction aux classes d'objet », *Langages* 131, Paris : Larousse, pp. 6-33.

LE PESANT, Denis. 2000a. « Compte rendu du figement lexical, par Salah Mejri ». *Linguisticae Investigationes* XXIII : 1, Amsterdam-Philadelphia : John Benjamins Publishing Co.

LE PESANT, Denis. 2000b. *Six études de sémantique lexicale sur les noms communs de lieux*, Habilitation à Diriger des Recherches, Paris : Université Paris XIII.

LE PESANT, Denis. 2001. « Les relations lexicales dans le vocabulaire des noms locatifs ». AUPELF-UREF, In *Actes des sixièmes journées LTT*, Beyrouth.

LE PESANT, Denis. 2002. « La détermination dans les anaphores ». *Langages* 145, Paris : Larousse.

LERAT, Pierre. 1995. *Les langues spécialisées*. Paris : PUF.

MARINIS, Theodore. 1998. « The acquisition of expletive definite articles in Modern Greek », In *Proceedings of ConSOLE 6, Lisbonne (décembre 1997)*, T. Cambier-Langeveld ; A. Lipták ; M. Radford (éds.), pp. 169-184.

MARTIN, Robert. 1997. « Sur les facteurs du figement lexical », *La locution entre langue et usages*, M. Martins-Baltar (éd.), Paris : ENS éditions, Fontenay/Saint-Cloud, pp. 291-305.

MARTINET, André. 1967. « Syntagme et synthème », *La linguistique*, 2, Paris : PUF, pp. 1-14.

MARTINEAU, Claude ; Elsa TOLONE ; Stavroula VOYATZI. 2007. « Les Entintés Nommées : usage et degrés de précision et de désambiguïsation », in *Proceedings of the 26th conference on Lexis and Grammar*, Bonifacio.

MARTINS-BALTAR, Michel. (éd.). 1997. *La locution entre langue et usages*, vol. 3, Paris : ENS éditions, Fontenay/Saint-Cloud.

MATHIEU-COLAS, Michel. 1990. « Orthographe et Informatique: établissement d'un dictionnaire électronique des variantes graphiques ». *Langue Française* 87, Paris : Larousse.

MATHIEU-COLAS, Michel. 1994. *Les mots à trait d'union. Problèmes de lexicographie informatique*, collection « Études de sémantique lexicale », Paris : Didier Erudition.

MATHIEU-COLAS, Michel. 1996. « Essai de typologie des noms composés français », In *Cahiers de Lexicologie* 69, Paris : Didier Erudition, pp. 71-125.

MAUREL, Denis. 2004. « Les mots inconnus sont-ils des noms propres ? », Septièmes Journées internationales d'Analyse statistique des Données Textuelles (JADT 2004), Louvain-la-Neuve, Belgique, 10-12 mars.

- MCDONALD, D. 1996. « Internal and External Evidence in the Identification and Semantic Categorisation of Proper Names », In Boguraev B. et Pustejovsky J. (éds), *Corpus processing for lexical acquisition* (Language, Speech and Communication), Cambridge, London : MIT Press, pp. 21-37.
- MEJRI, Salah. 1997a. *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Thèse d'Etat, Tunis : Publications de la faculté des lettres de la Manouba.
- MEJRI, Salah. 1997b. « Défigement et jeux de mots », In *Etudes linguistiques* 3, Tunis : Publication de l'Association tunisienne de linguistique.
- MEJRI, Salah. 2003. « Le figement lexical », In *Cahiers de Lexicologie* 82 : 1, pp. 23-39.
- MEUNIER, Annie. 1977. « Sur les bases syntaxiques de la morphologie dérivationnelle », *Linguisticae Investigationes* 1 : 2 ; Amsterdam/Philadelphia : John Benjamins, pp. 287-332.
- MEUNIER, Annie. 1981. *Nominalisation d'adjectifs par verbes supports*, Thèse de 3^{ème} cycle, Paris : Université Paris VII, Paris : LADL.
- MICHAILIDIS, Ionas ; Konstantinos, DIAMANTARAS ; Spyros VASILEIADIS ; Yannick, FRERE. 2006. « Greek Named Entity Recognition using Support Vector Machines, Maximum Entropy and Onetime », In *Proceedings of the LREC 2006 Conference*, Genoa, Italy.
- MIRAMBEL, André. 1953. « Les tendances actuelles de la dialectologie néohellénique », *Orbis* 2, Paris, pp. 448-472.
- MITSKA A. ; I. MICHAILIDIS. 2003. «Αυτόματη επεξεργασία αρκτικολέξων και ακρωνύμων νέας ελληνικής», In *Actes de la 24^{ème} Rencontre Annuelle de la Section de Linguistique du Département de Philologie de l'Université Aristote de Thessalonique*, pp. 482-493.
- MONCEAUX, Anne. 1993. *La formation des noms composés de structure Nom Adjectif. Elaboration d'un lexique électronique*, Thèse de doctorat, Paris : Université Paris VII.
- MOTA, C. ; P. CARVALHO ; E. Ranchhod. 2004. "Multiword Lexical Acquisition and Dictionary Formalization". In Michael Zock (ed.), *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries, COLING*, Geneva, pp. 73-76.

MOUSTAKI, Argyro. 1995. *Les expressions figées είμαι/être Prép C W en grec moderne*, Thèse de Doctorat, Université Paris VIII, Septentrion, Paris.

NAKAMURA, Takuya. 2005. « Analysing texts in a specific domain with local grammars : The case of stock exchange market reports », In *Linguistic Informatics - State of the Art and the Future*, Y. Kawaguchi et al. (éds.), Tokyo University of Foreign Studies, UBLI 1, Amsterdam/Philadelphia : Benjamins, pp. 76–98.

NEGRONI-PEYRE, Dominique de. 1978. « Nominalisation par être en et réflexivation (admiration, opposition, révolte et rage) », *Lingvisticae Investigationes* 2 : 1, Amsterdam : John Benjamins B.V, pp. 127-164.

PANAYOTOPOULOU, Styliani. 1992. *Les noms composés du type Adjectif Nom (AN) en grec moderne*, Mémoire de D.E.A., Paris : Université Paris VIII.

PANTAZARA, Andromaque-Virginie. 2003. *Syntaxe dérivationnelle du grec moderne : les constructions verbales à un complément prépositionnel et les constructions nominales et adjectivales prédicatives associées*, Thèse de doctorat, Paris : Université Paris VIII.

PAPANASTASIOU, Kostas. 2001. « Ορθοέπεια 5 : Μεταβολές και επιδράσεις του τελικού – v », In *ΟΡΟΓΡΑΜΜΑ* 48 (revue en ligne), pp.3-4. (cf. <http://sfr.ee.teiath.gr/htmSELIDES/Technology/Orogramma/Orogr-48d.htm>)

PAUMIER, Sébastien. 2003. *De la reconnaissance de formes linguistiques à l'analyse syntaxique*, Thèse de Doctorat, Paris : Université de Marne-la-Vallée.

PAUMIER, Sébastien. 2006. *Unitex 1.2 Manuel d'utilisation*, Université de Marne-la-Vallée. (cf. <http://www-igm.univ-mlv.fr/~unitex>)

PERRIN, Dominique. 1989. « Automates et algorithmes sur les mots », In *Annales des Télécommunications* 44 : 1-2, pp. 35-51.

PHILIPPAKI-WARBURTON, Irène. 1982. « Προβλήματα σχετικά με τη σειρά των όρων στις ελληνικές προτάσεις », In *Glossologia* 1, Athènes.

POIBEAU Thierry, 2005. « Sur le statut référentiel des entités nommées », In *Actes de la conférence Traitement Automatique des Langues Naturelles (TALN 2005)*, Dourdan. France.

PONCET-MONTANGE, Anne. 1991. *Les groupes nominaux de structure NAN et NAV*, Thèse de doctorat, Paris : Université Paris XIII.

RALLI, Angeliki. 2004. *Μορφολογία*, Athènes : Patakis.

RANCHHOD, E. 2005. "Using Corpora to Increase Portuguese MWE Dictionaries. Tagging MWE in a Portuguese Corpus". In: *Proceedings from The Corpus Linguistics Conference Series, Vol. 1, no. 1, Corpus Linguistics 2005*.

ROCHE, Emmanuel. 1992. « Text disambiguation by finite state automata, an algorithm and experiments on corpora », In *Proceedings of COLING-92*, Nantes.

ROCHE, Emmanuel. 1993. *Analyse syntaxique transformationnelle du français par transducteurs et lexique-grammaire*, Thèse de doctorat, Paris : Université Paris VII.

ROCHE, Emmanuel ; Y. SCHABES. 1997. "Parsing with Finite-State Transducers", In *Finite-State Language Processing*, E. Roche ; Y. Schabes (éds.), Cambridge, Mass./London, MIT Press.

SALKOFF, Morris. 1973. *Une grammaire en chaîne du français. Analyse distributionnelle*, Paris : Dunod.

SALKOFF, Morris. 1983. "Bees are swarming in the garden", *Language* 59 : 2, Baltimore : The Waverly Press, pp. 288-346.

SALKOFF, Morris. 1999. « A study of ambiguity using INTEX », *Linguisticae Investigationes* 22, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 143-154.

SAVARY, Agata. 2009. "Multiflex: a Multilingual Finite-State Tool for Multi-Word Units", in *Proceedings of CIAA 2009*, Sydney, Australie.

SCHAPIRA, Charlotte. 2000. *Les stéréotypes en français. Proverbes et autres formules*, Paris : Orphys.

SEKINE, S. ; C. NOBATA. 1998. « An Information Extraction System and a Customization Tool », in *Proceedings of the New Challenges in Natural Language Processing and its Application*, 25-26 May 1998, Tokyo, Japan.

SEKINE, S. ; K. SUDO ; C. NOBATA. 2002. « Extended Named Entity Hierarchy », in *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Canary Islands, Spain, pp. 1818-1824.

SENELART, Jean. 1999. *Outils de reconnaissance d'expressions linguistiques complexes dans de grands corpus*, Thèse de doctorat, Paris : Université Paris VII, Paris : LADL.

SETATOS, Michalis. 1992. « Συσχετιστές, τελεστές και ενδείκτες της κοινής νέας ελληνικής », *Studies in Greek Linguistics, Proceedings of the 13th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, (7-9 May 1992), Thessaloniki, pp. 351-369.

SFETSIU, Vasso. 2002. « Les constructions nominales à verbe support en grec moderne », *Recherches en Linguistique Grecque II : Actes du 5ème Colloque International de Linguistique Grecque, Sorbonne (13-15 septembre 2001)*, Paris : L'Harmattan, pp. 219-221.

SFETSIU, Vassileia. 2007. *Κατηγορηματικά ονόματα: Μέθοδος ανάλυσής τους για ηλεκτρονικές εφαρμογές*, Thèse de doctorat, Université Aristote de Thessaloniki.

SILBERZTEIN, Max. 1990. « Le dictionnaire électronique des mots composés », *Langue Française* 87, Paris : Larousse.

SILBERZTEIN, Max. 1993a. *Dictionnaires électroniques et analyse automatique de textes : le système INTEX*, Paris : Masson.

SILBERZTEIN, Max. 1993b. « Les groupes nominaux productifs et les noms composés lexicalisés », *Linguisticae Investigationes* 18 : 2, Amsterdam-Philadelphia : John Benjamins Publishing Co, pp. 405-425.

SILBERZTEIN, Max. 1997. « The lexical analysis of natural languages », In *Finite –state language processing*, E. Roche ; Y. Schabès (éds.), Cambridge, Mass./London : MIT Press.

SILBERZTEIN, Max. 1998. « Normalisation des textes », In *Actes des Troisièmes Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 1998)*, Nice, pp. 601-614.

SILBERZTEIN, Max. 1999. « INTEX : a Finite State Transducer Toolbox », In *Theoretical Computer Science* 231 : 1, pp. 33-46.

SKLAVOUNOU, Elsa. 1993. *Un Lexique-Grammaire trilingue de noms composés (grec-français-anglais) – Application au vocabulaire spécialisé du tennis*, D.E.A., 2 vol., Paris : Université Paris VIII.

SKLAVOUNOU, Elsa. 1997. *Etude comparée de la nominalisation des adjectifs en grec moderne et en français*, Thèse de Doctorat, Paris : Université Paris VIII.

SKLAVOUNOU, Elsa. 1998. « Dictionnaire morphologique électronique des adjectifs et des substantifs du grec moderne », *Studies in Greek Linguistics, Proceedings of the 19th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, (23-25 april 1998), Thessaloniki.

THOMADAKI, Evangelia. 1996. « Παράγωγα στη Νέα Ελληνική », In *Ζητήματα Νεοελληνικής Γλώσσας-Διδακτική προσέγγιση*, G. Katsimali ; F. Kavoukopoulos (éds.), Rethymno : Université de Crète.

TRIANTAPHYLLIDIS, Manolis. 1941. *Νεοελληνική Γραμματική της δημοτικής*, ΟΕΣΒ, Réimpression 2005 : Thessaloniki : Institut des Etudes Néohelléniques, [Fondation Manolis Triantaphyllidis].

TSAKNAKI, Olympia. 2005. *Problématique du proverbe dans la traduction : son usage en grec moderne et son traitement automatique*, Thèse de Doctorat, Thessaloniki : Université Aristote de Thessaloniki.

TSAMADOU, Irène. 1984. *Le génitif en grec moderne*, Thèse de doctorat, Paris : Université Paris VII.

TSAMADOU-JACOBBERGER, Irène. 1989. « Le déterminant 0 en grec moderne. Etude de quelques constructions. », *Mémoire du CERIL : Séminaire de Linguistique Grecque 1987-88*, vol. 4, Paris : Université Paris VII, pp. 69-81.

TSAMADOU-JACOBBERGER, Irène. 1990. « Ο δείκτης ΣΑΝ », *Studies in Greek Linguistics, Proceedings of the 12th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki*, 9-11 May 1989, Thessaloniki, pp. 295-306.

TSAMADOU-JACOBBERGER, Irène. 1993. « Etude comparée du déterminant zéro et de l'indéfini en grec et en français », In *Linguistique contrastive et Traduction*, Paris : Orphys, pp. 115-153.

TSAMADOU-JACOBBERGER, Irène. 2002. *Le nom en grec moderne : marqueurs et opérations de détermination*, Paris : L'Harmattan.

TSIMPLI, I. M. ; S. STAVRAKAKI. 1999. « The effects of a morphosyntactic deficit in the determiner system : The case of a Greek SLI child », In *Lingua* 108, pp. 31-85.

TSOPANAKIS, Agapitos. 1994. *Νεοελληνική Γραμματική*, Thessaloniki : Αφοί Κυριακίδη.

TZARTZANOS, Ahileas. 1946. *Νεοελληνική σύνταξις (της κοινής δημοτικής). Τόμοι Α', Β'*, Thessaloniki : Αφοί Κυριακίδη (1989).

TZARTZANOS, Ahileas. 1954. (2^{ème} éd.). *Γραμματική της νέας ελληνικής γλώσσας (της απλής καθαρευούσης)*, Αθήνα : Κ. Κακουλίδης.

TZARTZANOS, Ahileas. 1967. *Συντακτικό της αρχαίας ελληνικής γλώσσας*, Thessaloniki : Afoi Kyriakidi.

VASSILAKI, Sofia ; Irène, TSAMADOU-JACOBBERGER. 1993. « Οι επιφωνηματικές προτάσεις στα Νέα Ελληνικά », *Studies in Greek Linguistics, Proceedings of the 14th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki, 27-29 April 1993*, Thessaloniki, pp. 98-112.

VOYATZI, Stavroula. 2004. « Représentation électronique et reconnaissance automatique des déterminants numériques du grec moderne », In *Studies in Greek Linguistics, Proceedings of the 24th Annual Meeting of the Department of Linguistics, Faculty of Philology, Aristotle University of Thessaloniki, 9-11 mai 2003*, Thessaloniki, pp. 78-89 (en grec).

VOYATZI, Stavroula. 2006. *Description morphosyntaxique et sémantique des adverbes figés en vue d'un système d'analyse automatique des textes grecs*, Thèse de doctorat, Paris : Université de Marne-la-Vallée.

WATRIN, Patrick. 2006. *Une approche hybride de l'extraction d'information : sous-langages et lexique-grammaire*, Thèse de doctorat, Louvain-la-Neuve : Cental, Université de Louvain-la-Neuve.

Dictionnaires

ANASTASSIADIS-SYMEONIDIS, Anna. 2002. *Αντίστροφο Λεξικό της Νέας Ελληνικής*, Thessaloniki : Insitut des Etudes Néohelléniques, [Fondation Manolis Triantaphyllidis].

BABINIOTIS, Georgios. 1998. *Λεξικό της νέας ελληνικής γλώσσας*, Athènes : Centre de Lexicologie.

Insitut des Etudes Néohelléniques, [Fondation Manolis Triantaphyllidis]. 1998. *Λεξικό της Κοινής Νεοελληνικής*, Thessaloniki.